OM protein - protein search, using sw model

```
Run on:          May 26, 2004, 18:23:35 ; Search time 0.852781 Seconds
                                           (without alignments)
                                           3496.721 Million cell updates/sec

Title:           US-09-302-812-11
Perfect score:   167
Sequence:        1 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 31

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_78:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	47.3	997	2	B84726	probable poly(ADP-
2	53	31.7	726	2	T21138	hypothetical prote
3	52	31.1	510	1	S43516	carboxypeptidase C
4	51	30.5	132	2	B70008	hypothetical prote
5	50.5	30.2	541	2	G64441	hypothetical prote
6	50	29.9	166	2	AB0371	conserved hypothet
7	50	29.9	364	2	T41929	polymerase process
8	50	29.9	364	2	A84726	probable poly(ADP-
9	50	29.9	456	2	A85139	hypothetical prote
10	50	29.9	499	1	CPBHS	carboxypeptidase C
11	49.5	29.6	1213	2	A58198	serine/proline-ric
12	49	29.3	152	2	I47188	Ig heavy chain var
13	48.5	29.0	558	2	C96933	2-isopropylmalate
14	48	28.7	250	2	A69528	2-hydroxyhepta-2,4
15	48	28.7	474	2	T10271	capsid-associated
16	48	28.7	521	1	DNYC27	NADH2 dehydrogenas
17	47.5	28.4	844	1	QQVZ1	mRNA capping enzym
18	47.5	28.4	844	1	QQVZVC	mRNA capping enzym
19	47.5	28.4	844	2	T28529	probable mRNA guan
20	47.5	28.4	844	2	A72162	F1R protein - vari
21	47.5	28.4	844	2	T37374	mRNA capping enzym
22	47.5	28.4	1745	2	S44816	F44E2.1 protein -
23	47.5	28.4	2272	2	T18572	gag, pol and env p
24	47	28.1	150	2	A42282	ferric uptake regu
25	47	28.1	459	2	AD3223	hypothetical prote
26	47	28.1	488	2	E95382	hypothetical prote
27	47	28.1	520	2	AC2416	NADH dehydrogenase
28	47	28.1	670	2	AI1847	two-component sens
29	47	28.1	885	2	T26310	hypothetical prote
30	46.5	27.8	164	2	B47207	phycoerythrin alph
31	46.5	27.8	164	2	S73289	phycoerythrin alph
32	46.5	27.8	304	2	AD0864	probable membrane
33	46.5	27.8	401	2	B89863	argininosuccinate
34	46.5	27.8	515	2	A43289	CDC4L protein - hu
35	46.5	27.8	796	2	T39962	coatomer complex b
36	46	27.5	77	2	A37317	probable membrane
37	46	27.5	396	2	AC1849	hypothetical prote
38	46	27.5	445	2	T26762	hypothetical prote
39	46	27.5	468	2	AG2899	dihydrolipoamide d
40	46	27.5	468	2	H97674	dihydrolipoamide d
41	46	27.5	491	2	AC1816	hypothetical prote
42	46	27.5	518	2	I39989	spore cortex synth
43	46	27.5	527	2	A86607	lysyl tRNA synthet
44	46	27.5	527	2	E72016	lysine-tRNA ligase
45	46	27.5	601	2	B75136	hypothetical prote

ALIGNMENTS

```
RESULT 1
B84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31870
A;Map position: 2

  Query Match             47.3%;  Score 79;  DB 2;  Length 997;
  Best Local Similarity   48.4%;  Pred. No. 0.0049;
  Matches   15;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 31
              ||   :| || : | | |::| ||||| ::|
Db        274 LFLPRMDDNEAIEIVGAERFSCYTGYASSFR 304

RESULT 2
T21138
hypothetical protein F20C5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T21138
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19381
A;Accession: T21138
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-726 <WIL>
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1
A;Experimental source: clone F20C5
C;Genetics:
A;Gene: CESP:F20C5.1
A;Map position: 4
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

  Query Match             31.7%;  Score 53;  DB 2;  Length 726;
```

```
  Best Local Similarity   35.5%;  Pred. No. 18;
  Matches   11;  Conservative    4;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 31
              |  | :   | : | |    :| ||||  | :
Db        552 LLCEKMKQLEAISIVGAYVFSSYTGYGHTLK 582
```

RESULT 3
S43516
carboxypeptidase C (EC 3.4.16.5) precursor - rice
N;Alternate names: serine-type carboxypeptidase I
C;Species: Oryza sativa (rice)
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 16-Jun-2000
C;Accession: S43516
R;Washio, K.; Ishikawa, K.
Biochim. Biophys. Acta 1199, 311-314, 1994
A;Title: Cloning and sequencing of the gene for type I carboxypeptidase in rice.
A;Reference number: S43516; MUID:94213891; PMID:8161571
A;Accession: S43516
A;Molecule type: DNA
A;Residues: 1-510 <WAS>
A;Cross-references: EMBL:D17586; NID:g409579; PIDN:BAA04510.1; PID:g409580
A;Note: the authors translated the codon AGT for residue 430 as Arg
C;Genetics:
A;Introns: 60/2; 109/1; 134/3; 175/3; 210/1; 222/3; 256/3; 283/3; 367/3; 389/3; 430/2; 4
C;Superfamily: serine carboxypeptidase
C;Keywords: disulfide bond; glycoprotein; hydrolase; serine carboxypeptidase
F;1-25/Domain: signal sequence #status predicted <SIG>
F;154,268,418/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;194,434,487/Active site: Ser, Asp, His #status predicted

```
  Query Match             31.1%;  Score 52;  DB 1;  Length 510;
  Best Local Similarity   45.0%;  Pred. No. 17;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          7 DHNECLIITGTEQYSEYTGY 26
              ||: |:  |||| ::   ||
Db        432 DHDMCVPYTGTEAWTRSLGY 451
```

RESULT 4
B70008
hypothetical protein yueI - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: B70008
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B70008
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-132 <KUN>
A;Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15165.1; PID:g2635672
A;Experimental source: strain 168
C;Genetics:
A;Gene: yueI
C;Superfamily: Bacillus subtilis hypothetical protein yueI

```
  Query Match             30.5%;  Score 51;  DB 2;  Length 132;
  Best Local Similarity   42.9%;  Pred. No. 5.7;
  Matches    9;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          8 HNECLIITGTEQYSEYTGYAE 28
              ||  |:| |  ||  |: | :
Db         62 HNVTLLINGELQYQSYSSYIQ 82
```

RESULT 5
G64441
hypothetical protein Lpg22p homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: G64441
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: G64441
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-541 <BUL>
A;Cross-references: GB:U67556; GB:L77117; NID:g1591766; PIDN:AAB99138.1; PID:g1591770; T
C;Genetics:
A;Map position: FOR1076255-1077880
A;Start codon: TTG
C;Superfamily: hypothetical protein YPL086c

```
  Query Match             30.2%;  Score 50.5;  DB 2;  Length 541;
  Best Local Similarity   30.2%;  Pred. No. 31;
  Matches   13;  Conservative    6;  Mismatches    9;  Indels   15;  Gaps    2;

Qy          1 LFTEVLDHNE----------CLIITGTEQY-----SEYTGYAE 28
              :| |: :: :          ||:| |||  |        ||  | |
Db        292 MFKEIFENPDFKPDMVKIYPCLVIEGTELYEMWKRGEYKPYRE 334
```

RESULT 6
AB0371
conserved hypothetical protein YPO3051 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AB0371
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0371
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-166 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92293.1; PID:g15981004; GSPDB:GN00175
C;Genetics:
A;Gene: YPO3051

```
  Query Match             29.9%;  Score 50;  DB 2;  Length 166;
  Best Local Similarity   33.3%;  Pred. No. 10;
  Matches   10;  Conservative    5;  Mismatches   11;  Indels    4;  Gaps    1;

Qy          2 FTEVLDHNECLIITGTEQYSEYTGYAETYR 31
              :   ::||    : |||  : | |   |||
Db         64 YRAAIEHN----LVGTELQAAYLGLGSTYR 89
```

RESULT 7
T41929
polymerase processivity factor - human herpesvirus 7 (strain JI)

```
C;Species: human herpesvirus 7
A;Variety: strain JI
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41929
R;Nicholas, J.
submitted to the EMBL Data Library, December 1995
A;Description: Determination and analysis of the complete nucleotide sequence of human h
A;Reference number: Z22022
A;Accession: T41929
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-364 <NIC>
A;Cross-references: EMBL:U43400; PIDN:AAC54689.1
A;Experimental source: strain JI
C;Genetics:
A;Note: U27
C;Superfamily: human herpesvirus 6 P41 protein

  Query Match             29.9%;  Score 50;  DB 2;  Length 364;
  Best Local Similarity   41.7%;  Pred. No. 24;
  Matches   10;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 LFTEVLDHNECLIITGTEQYSEYT 24
                |   ::: | ||| || |: :|   |
Db         70 LVLKLIIHAECLYITDTDHFSTKT 93


RESULT 8
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

  Query Match             29.9%;  Score 50;  DB 2;  Length 364;
  Best Local Similarity   36.0%;  Pred. No. 24;
  Matches    9;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 LFTEVLDHNECLIITGTEQYSEYTG 25
                ||    ::  |  : :  | |::| |||
Db        213 LFLPTMEVTEAIEVVGAERFSLYTG 237


RESULT 9
A85139
hypothetical protein AT4g12910 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C;Accession: A85139
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: A85139
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-456 <STO>
A;Cross-references: GB:NC_001268; NID:g7267993; PIDN:CAB78333.1; GSPDB:GN00140
C;Genetics:
```

```
A;Gene: AT4g12910
A;Map position: 4
C;Superfamily: serine carboxypeptidase

  Query Match             29.9%;  Score 50;  DB 2;  Length 456;
  Best Local Similarity   40.0%;  Pred. No. 30;
  Matches    8;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          7 DHNECLIITGTEQYSEYTGY 26
                ||: |:  ||:| :::  ||
Db        391 DHDMCVPFTGSEAWTKSLGY 410


RESULT 10
CPBHS
carboxypeptidase C (EC 3.4.16.5) precursor - barley
N;Alternate names: serine-type carboxypeptidase I
C;Species: Hordeum vulgare (barley)
C;Date: 30-Jun-1988 #sequence_revision 24-Nov-1999 #text_change 16-Jun-2000
C;Accession: T05367; A29226; A25858; B25858
R;Rocher, A.; Lok, F.; Cameron-Mills, V.; von Wettstein, D.
submitted to the EMBL Data Library, November 1996
A;Description: The gene family of serine carboxypeptidases in barley.
A;Reference number: Z15412
A;Accession: T05367
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-499 <ROC>
A;Cross-references: EMBL:Y09603; PIDN:CAA70816.1
A;Experimental source: cv. Himalaya, aleurone
R;Doan, N.P.; Fincher, G.B.
J. Biol. Chem. 263, 11106-11110, 1988
A;Title: The A- and B-chains of carboxypeptidase I from germinated barley originate from
A;Reference number: A29226; MUID:88298749; PMID:3403516
A;Accession: A29226
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 88-499 <DOA>
A;Cross-references: GB:J03897; NID:g167011; PIDN:AAA32940.1; PID:g167012
R;Sorensen, S.B.; Breddam, K.; Svendsen, I.
Carlsberg Res. Commun. 51, 475-485, 1986
A;Title: Primary structure of carboxypeptidase I from malted barley.
A;Reference number: A90769
A;Accession: A25858
A;Molecule type: protein
A;Residues: 31-101,'P',103-296;352-424,'X',426-499 <SOR>
C;Genetics:
A;Gene: Cxp1
C;Complex: homodimer of two mature products, each containing an A and a B chain
C;Superfamily: serine carboxypeptidase
C;Keywords: disulfide bond; glycoprotein; homodimer; hydrolase; serine carboxypeptidase;
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-296,352-499/Product: carboxypeptidase C #status experimental <MAT>
F;31-296/Domain: chain A #status experimental <CHA>
F;297-351/Domain: connecting peptide #status predicted <CNP>
F;352-499/Domain: chain B #status experimental <CHB>
F;148,262,407/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;188,423,476/Active site: Ser, Asp, His #status predicted

  Query Match             29.9%;  Score 50;  DB 1;  Length 499;
  Best Local Similarity   40.0%;  Pred. No. 33;
  Matches    8;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          7 DHNECLIITGTEQYSEYTGY 26
                ||: |:  ||:| :::  ||
Db        421 DHDMCVPFTGSEAWTKSLGY 440


RESULT 11
A58198
serine/proline-rich FEL protein, splice form 1 - human
N;Alternate names: AF-4 protein
C;Species: Homo sapiens (man)
```

```
C;Date: 02-Jul-1996 #sequence_revision 25-Oct-1996 #text_change 01-Dec-2000
C;Accession: A58198; I52572
R;Morrissey, J.J.M.
Blood 81, 1124-1131, 1993
A;Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.
A;Reference number: I52572; MUID:93184301; PMID:8443374
A;Accession: A58198
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1213 <MOR>
A;Cross-references: GB:L25050; NID:g407323
A;Accession: I52572
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'MAAQS',13-1213 <MO2>
A;Cross-references: GB:L25050; NID:g407323; PIDN:AAA36642.1; PID:g407324
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit
C;Genetics:
A;Gene: GDB:MLLT2; AF-4; FEL
A;Cross-references: GDB:136792; OMIM:159557
A;Map position: 4q21-4q21
C;Keywords: alternative splicing

  Query Match             29.6%;  Score 49.5;  DB 2;  Length 1213;
  Best Local Similarity   46.4%;  Pred. No. 1e+02;
  Matches   13;  Conservative    3;  Mismatches    9;  Indels    3;  Gaps    1;

Qy          5 VLDHNECLIITGTEQYSE---YTGYAET 29
              ||  ||| | | :|  |     |: |:||
Db       1002 VLSFNECGIATESESQSSKSAYSVYSET 1029

RESULT 12
I47188
Ig heavy chain variable VDJ region - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 21-Jan-2000
C;Accession: I47188
R;Sun, J.; Kacskovics, I.; Brown, W.R.; Butler, J.E.
J. Immunol. 153, 5618-5627, 1994
A;Title: Expressed swine VH genes belong to a small VH gene family homologous to human V
A;Reference number: I47177; MUID:95081609; PMID:7989761
A;Accession: I47188
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-152 <SUN>
A;Cross-references: EMBL:U15447; NID:g571380; PIDN:AAA67013.1; PID:g571381
C;Superfamily: immunoglobulin V region; immunoglobulin homology
F;34-119/Domain: immunoglobulin homology <IMM>

  Query Match             29.3%;  Score 49;  DB 2;  Length 152;
  Best Local Similarity   40.9%;  Pred. No. 13;
  Matches    9;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         10 ECLIITGTEQYSEYTGYAETYR 31
              |||    |:  || :| ||:: :
Db         65 ECLASIGSSSYSGHTYYADSVK 86

RESULT 13
C96933
2-isopropylmalate synthase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: C96933
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C96933
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 1-558 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78254.1; PID:g15023111; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0273
C;Superfamily: 2-isopropylmalate synthase

  Query Match             29.0%;  Score 48.5;  DB 2;  Length 558;
  Best Local Similarity   37.5%;  Pred. No. 61;
  Matches   12;  Conservative    7;  Mismatches   10;  Indels    3;  Gaps    1;

Qy          1 LFTEVLDHNECLIITGTE---QYSEYTGYAET 29
              :| :   |   :  | | | :    :||||| |:|:
Db        136 VFKKDKDEVKSLAIKGAQMVKKYSEYTEYSES 167

RESULT 14
A69528
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-2) homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: A69528
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodsor
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaec
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: A69528
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-250 <KLE>
A;Cross-references: GB:AE000952; GB:AE000782; NID:g2689275; PIDN:AAB89031.1; PID:g264830
C;Superfamily: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

  Query Match             28.7%;  Score 48;  DB 2;  Length 250;
  Best Local Similarity   34.8%;  Pred. No. 31;
  Matches    8;  Conservative    6;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 TEVLDHNECLIITGTEQYSEYTG 25
              | |: |::|:|:      |  :| |
Db         76 TAVIGHDDCIILPQISQRVDYEG 98

RESULT 15
T10271
capsid-associated protein 1629 - Orgyia pseudotsugata nuclear polyhedrosis virus
C;Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNPV
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C;Accession: T10271
R;Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.
Virology 229, 381-399, 1997
A;Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis
A;Reference number: Z17011; MUID:97271300; PMID:9126251
A;Accession: T10271
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-474 <AHR>
A;Cross-references: EMBL:U75930; NID:g2934903; PID:g2934904

  Query Match             28.7%;  Score 48;  DB 2;  Length 474;
  Best Local Similarity   31.2%;  Pred. No. 60;
  Matches   10;  Conservative   10;  Mismatches   10;  Indels    2;  Gaps    1;

Qy          1 LFTEVLDHNECLIITGT--EQYSEYTGYAETY 30
              |   :|::  |   :::|    |:||:|   | :|:
Db        116 LRVDVVNDAEVNVLSGDFYEEYSKYISYQQTF 147

Search completed: May 26, 2004, 18:47:57
Job time : 1.85278 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          May 26, 2004, 17:19:25 ; Search time 111.936 Seconds
                                          (without alignments)
                                          2466.140 Million cell updates/sec

Title:           US-09-302-812-2
Perfect score:   5184
Sequence:        1 MSAGPGCEPCTKRPRWDAAA..........YHAVESCTQTTNQPGQRTGA 977

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
                 1:  geneseqp1980s:*
                 2:  geneseqp1990s:*
                 3:  geneseqp2000s:*
                 4:  geneseqp2001s:*
                 5:  geneseqp2002s:*
                 6:  geneseqp2003as:*
                 7:  geneseqp2003bs:*
                 8:  geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5184	100.0	977	5	AAE25629	Aae25629 Bovine po
2	5184	100.0	977	5	AAU76020	Aau76020 Bovine po
3	5184	100.0	977	5	AAU75799	Aau75799 Bovine po
4	5184	100.0	977	6	ABG72278	Abg72278 Bovine po
5	4635.5	89.4	976	5	AAE25630	Aae25630 Human pol
6	4635.5	89.4	976	5	AAU76021	Aau76021 Human pol
7	4635.5	89.4	976	5	AAU76012	Aau76012 Human pol
8	4635.5	89.4	976	6	ABG72279	Abg72279 Human pol
9	4332.5	83.6	968	5	AAE25631	Aae25631 Murine po
10	4332.5	83.6	968	5	AAU76022	Aau76022 Mouse pol
11	4332.5	83.6	968	5	AAU76013	Aau76013 Mouse pol
12	4332.5	83.6	968	6	ABG72280	Abg72280 Mmurine p
13	1063.5	20.5	768	4	ABB59491	Abb59491 Drosophil
14	1063.5	20.5	768	5	AAE25632	Aae25632 Fruit fly
15	1063.5	20.5	768	5	AAU76023	Aau76023 Fruit fly
16	1063.5	20.5	768	5	AAU76014	Aau76014 Fruit fly
17	1063.5	20.5	768	6	ABG72281	Abg72281 Fruit fly
18	465	9.0	726	5	AAE25633	Aae25633 Poly aden
19	465	9.0	726	5	AAU76024	Aau76024 Worm poly
20	465	9.0	726	5	AAU76015	Aau76015 Worm poly
21	465	9.0	726	6	ABG72282	Abg72282 C. elegan
22	335.5	6.5	819	4	ABG20721	Abg20721 Novel hum
23	201	3.9	100	4	ABG11103	Abg11103 Novel hum
24	178	3.4	33	5	AAE25652	Aae25652 Bovine po
25	174.5	3.4	2194	4	AAM40114	Aam40114 Human pol
26	174	3.4	1046	4	ABG10125	Abg10125 Novel hum
27	170	3.3	1284	4	ABG09663	Abg09663 Novel hum
28	170	3.3	1284	4	ABG18482	Abg18482 Novel hum
29	170	3.3	1284	4	ABG06083	Abg06083 Novel hum
30	170	3.3	1284	4	ABG10823	Abg10823 Novel hum
31	170	3.3	1300	4	ABG09142	Abg09142 Novel hum
32	170	3.3	1308	4	ABG04003	Abg04003 Novel hum
33	167	3.2	31	5	AAE25634	Aae25634 Bovine po
34	167	3.2	31	5	AAE25651	Aae25651 Bovine po
35	167	3.2	31	5	AAU76025	Aau76025 Bovine po
36	167	3.2	31	5	AAU76016	Aau76016 Bovine po
37	167	3.2	31	6	ABG72283	Abg72283 Oligopept
38	166	3.2	962	4	ABG04842	Abg04842 Novel hum
39	166	3.2	1462	6	ABP58346	Abp58346 Human cel
40	166	3.2	1462	6	ABU05132	Abu05132 Human exp
41	166	3.2	1462	6	ABU05136	Abu05136 Human exp
42	166	3.2	1462	6	ABU05131	Abu05131 Human exp
43	166	3.2	1462	6	ABU05135	Abu05135 Human exp
44	166	3.2	1524	4	ABG04845	Abg04845 Novel hum
45	164.5	3.2	1278	4	ABB71882	Abb71882 Drosophil

ALIGNMENTS

```
RESULT 1
AAE25629
ID   AAE25629 standard; protein; 977 AA.
XX
AC   AAE25629;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).
XX
KW   Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Bos taurus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42081.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 47-45; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
```

```
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is bPARG
XX
SQ   Sequence 977 AA;

  Query Match             100.0%;  Score 5184;  DB 5;  Length 977;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  977;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        961 VESCTQTTNQPGQRTGA 977
                |||||||||||||||||
Db        961 VESCTQTTNQPGQRTGA 977

RESULT 2
AAU76020
ID   AAU76020 standard; protein; 977 AA.
XX
AC   AAU76020;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Region          601..617
FT                   /note= "Represents PARG oligopeptide #1"
FT   Region          761..770
FT                   /note= "Represents PARG oligopeptide #2"
FT   Region          771..801
FT                   /note= "Represents PARG oligopeptide #3"
FT   Region          849..880
FT                   /note= "Represents PARG oligopeptide #4"
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14931.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 47-52; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
```

```
CC   disorders. The present amino acid sequence represents the bovine PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 977 AA;

  Query Match             100.0%;  Score 5184;  DB 5;  Length 977;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  977;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        961 VESCTQTTNQPGQRTGA 977
              |||||||||||||||||
Db        961 VESCTQTTNQPGQRTGA 977

RESULT 3
AAU75799
ID   AAU75799 standard; protein; 977 AA.
XX
AC   AAU75799;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Region          601..617
FT                   /note= "Represents PARG oligopeptide #1"
FT   Region          761..770
FT                   /note= "Represents PARG oligopeptide #2"
FT   Region          771..801
FT                   /note= "Represents PARG oligopeptide #3"
FT   Region          849..880
FT                   /note= "Represents PARG oligopeptide #4"
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14493.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 45-52; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the bovine PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 977 AA;

  Query Match             100.0%;  Score 5184;  DB 5;  Length 977;
  Best Local Similarity   100.0%;  Pred. No. 0;
```

```
Matches  977;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        961 VESCTQTTNQPGQRTGA 977
              |||||||||||||||||
Db        961 VESCTQTTNQPGQRTGA 977

RESULT 4
ABG72278
```

```
ID   ABG72278 standard; protein; 977 AA.
XX
AC   ABG72278;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Bos taurus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14477.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents bovine PARG enzyme
XX
SQ   Sequence 977 AA;

  Query Match             100.0%;  Score 5184;  DB 6;  Length 977;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  977;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy         121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy         181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy         241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy         301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy         361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy         421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy         481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy         541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy         601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy         661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy         721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy         781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy         841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy         901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy         961 VESCTQTTNQPGQRTGA 977
               |||||||||||||||||
Db         961 VESCTQTTNQPGQRTGA 977
```

```
RESULT 5
AAE25630
ID   AAE25630 standard; protein; 976 AA.
XX
AC   AAE25630;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Human poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Homo sapiens.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42082.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 55-60; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is human PARG
XX
SQ   Sequence 976 AA;

  Query Match             89.4%;  Score 4635.5;  DB 5;  Length 976;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy           1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
               |:|||||||||  || || ||  |||||||||| |||||| ||| |||||||||   |
Db           1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy          61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
               |:|||||||||||||||||||||||||||| ||||| ||||||| ||||||||||||
Db          60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy         121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
               |||:||| |||| |||  || |||| |||||||||||||| |:|:|||||  |||||
Db         120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy         181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
               |||||||:|:|  ||||||:|||:|||||||| |||||||||| |||  |||:||||| |
Db         180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy         241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
               :||| | || |||:| ||| :||||| ||||||| || |:| |||| |||||||||||
Db         240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
```

```
Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
              |||||||||||||||||||||||||||||||| ||:||||||||| |:|| | ||| |
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
              ||||:||||||||||| ||||:||| ||: |||||||||||||:|||||||| ||:||
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
              |||:|||| | |||||||||||:|||||||||||||||||||||||||||||| ||||||
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
              |||||||||| |||||||||:||||||||||||||||||||||||||||||| |||||||
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
              |||||: |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
              ||||||| ||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
              ||||||||||||||||||||||:||||| ||||| :|| ||||||||||||||||||||
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
              |:||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
              ||||||||||||||| |||||||||||:||||||||||||||||||||||||||||||||
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        961 VESCTQTTNQPGQRTG 976
              ||||  :|  :   |||||
Db        960 VESCAETADHSGQRTG 975
```

RESULT 6
AAU76021

```
ID   AAU76021 standard; protein; 976 AA.
XX
AC   AAU76021;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Homo sapiens.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14932.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 55-60; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the human PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 976 AA;

  Query Match             89.4%;  Score 4635.5;  DB 5;  Length 976;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              |:|||||||||   || || ||  |||||||||| |||||| ||| ||||||||   |
Db          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
              |:||||||||||||||||||||||||||||| |||||| ||||||||| ||||||||||||
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
              |||:||| |||||  |||  ||  || ||||||||||||||| |:|:|||||  |||||
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
              ||||||:|:|  ||||||:|||:||||||| |||||||||| |||  |||:||||| |
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
              :||| | || |||:| ||| :||||| |||||||| || |:| |||| ||||||||||||
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
              |||||||||||||||||||||||||||||||| ||:||||||||| |:|| | ||| |
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
              ||||:||||||||||| ||||:||| ||: |||||||||||||:|||||||| ||:||
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
              |||:|||| | |||||||||||:|||||||||||||||||||||||||||||| ||||||
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
              |||||||||| |||||||||:||||||||||||||||||||||||||||||| |||||||
```

```
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        961 VESCTQTTNQPGQRTG 976
Db        960 VESCAETADHSGQRTG 975
```

RESULT 7
AAU76012

```
ID   AAU76012 standard; protein; 976 AA.
XX
AC   AAU76012;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Homo sapiens.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14494.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 55-60; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the human PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 976 AA;

  Query Match             89.4%;  Score 4635.5;  DB 5;  Length 976;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
```

```
Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        961 VESCTQTTNQPGQRTG 976
Db        960 VESCAETADHSGQRTG 975

RESULT 8
ABG72279
ID   ABG72279 standard; protein; 976 AA.
XX
AC   ABG72279;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14478.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents human PARG enzyme
XX
SQ   Sequence 976 AA;

  Query Match             89.4%;  Score 4635.5;  DB 6;  Length 976;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
```

```
Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        961 VESCTQTTNQPGQRTG 976
Db        960 VESCAETADHSGQRTG 975
```

RESULT 9
AAE25631
ID   AAE25631 standard; protein; 968 AA.
XX
AC   AAE25631;
XX
DT   04-NOV-2002 (first entry)
XX
DE   Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Mus musculus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42083.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 63-68; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is murine PARG
XX
SQ   Sequence 968 AA;

```
  Query Match             83.6%;  Score 4332.5;  DB 5;  Length 968;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        961 VESCTQTTNQPGQRTG 976
Db        952 VESSAETTDMPGQKAG 967
```

RESULT 10
AAU76022
ID   AAU76022 standard; protein; 968 AA.
XX
AC   AAU76022;

```
XX
DT   08-MAY-2002  (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Mus musculus.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14933.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 63-70; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the mouse PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 968 AA;

  Query Match             83.6%;  Score 4332.5;  DB 5;  Length 968;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        961 VESCTQTTNQPGQRTG 976
Db        952 VESSAETTDMPGQKAG 967


RESULT 11
AAU76013
ID   AAU76013 standard; protein; 968 AA.
XX
AC   AAU76013;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Mus musculus.
XX
PN   US6333148-B1.
```

```
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14495.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 63-68; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the mouse PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 968 AA;

  Query Match             83.6%;  Score 4332.5;  DB 5;  Length 968;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        961 VESCTQTTNQPGQRTG 976
Db        952 VESSAETTDMPGQKAG 967

RESULT 12
ABG72280
ID   ABG72280 standard; protein; 968 AA.
XX
AC   ABG72280;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;
KW   cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;
KW   vasotropic; anticonvulsant; cerebroprotective; enzyme.
XX
OS   Mus musculus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
```

```
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14479.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents murine PARG enzyme
XX
SQ   Sequence 968 AA;

  Query Match             83.6%;  Score 4332.5;  DB 6;  Length 968;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
Db          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        961 VESCTQTTNQPGQRTG 976
Db        952 VESSAETTDMPGQKAG 967


RESULT 13
ABB59491
ID   ABB59491 standard; protein; 768 AA.
XX
AC   ABB59491;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5265.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03594.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
```

```
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 768 AA;

  Query Match             20.5%;  Score 1063.5;  DB 4;  Length 768;
  Best Local Similarity   41.5%;  Pred. No. 6.2e-77;
  Matches  244;  Conservative   93;  Mismatches  198;  Indels   53;  Gaps   15;

Qy        417 RVPKAEDKRKEQCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 472
              |: |: |     : | : :              | :|  | |  | |  :| :   | |    |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        473 PPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERA 532
              ||:      | |   :    | | | | | |      ||::||::||: :: || |: :|
Db         95 PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        533 AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 592
                 |||:|: |||   :    : |: ||: ||   |   :| | ||      |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        593 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 651
               :|| ::::|| ||::    |:||||    | |:::||:||: |||||| |||||||  | |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        652 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 703
              ||||::||||||||::            |||| :  |||||  ||:      |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        704 EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761
              |    :| :     |      |||  |||||  | |:||||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTE 821
              | |||:| :||||  |     | |:: | |:|| ||||| :: |: : || : ||      || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        822 IVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 875
              ||||||||  :    |:  : : |||||| ||:           ||           |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        876 FGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLR 935
              ||||: ||||:|::  |     |  : |:|||: |     | : |            || ::: :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        936 YYNEECRNCSTPGP------DIKLYPFIYHAVESCTQTTNQPGQRTGA 977
                |:   :  |:  |               ||| ||     |      :   : ||:  |
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGASA 606

RESULT 14
AAE25632
ID   AAE25632 standard; protein; 768 AA.
XX
AC   AAE25632;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy; fruit fly.
XX
OS   Drosophila melanogaster.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42084.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Example 13; Col 71-76; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is fruit fly PARG
XX
SQ   Sequence 768 AA;

  Query Match             20.5%;  Score 1063.5;  DB 5;  Length 768;
  Best Local Similarity   41.5%;  Pred. No. 6.2e-77;
  Matches  244;  Conservative   93;  Mismatches  198;  Indels   53;  Gaps   15;

Qy        417 RVPKAEDKRKEQCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 472
              |: |: |     : | : :              | :|  | |  | |  :| :   | |    |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        473 PPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERA 532
              ||:      | |   :    | | | | | |      ||::||::||: :: || |: :|
Db         95 PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        533 AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 592
                 |||:|: |||   :    : |: ||: ||   |   :| | ||      |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        593 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 651
               :|| ::::|| ||::    |:||||    | |:::||:||: |||||| |||||||  | |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        652 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 703
              ||||::||||||||::            |||| :  |||||  ||:      |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        704 EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761
              |    :| :     |      |||  |||||  | |:||||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTE 821
              | |||:| :||||  |     | |:: | |:|| ||||| :: |: : || : ||      || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        822 IVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 875
              ||||||||  :    |:  : : |||||| ||:           ||           |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        876 FGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLR 935
```

```
                 ||||: ||||:|:: |    | : |:|||: |    | : |            || ::: :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        936 YYNEECRNCSTPGP------DIKLYPFIYHAVESCTQTTNQPGQRTGA 977
              |:   :  |:  |          ||| ||    |   :  : ||:   |
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGASA 606

RESULT 15
AAU76023
ID   AAU76023 standard; protein; 768 AA.
XX
AC   AAU76023;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Fruit fly poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Fruit fly; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Drosophila melanogaster.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14934.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Example 13; Col 71-76; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the fruit fly PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 768 AA;

  Query Match             20.5%;  Score 1063.5;  DB 5;  Length 768;
  Best Local Similarity   41.5%;  Pred. No. 6.2e-77;
  Matches  244;  Conservative   93;  Mismatches  198;  Indels   53;  Gaps   15;

Qy        417 RVPKAEDKRKEQCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 472
              |: |: |      : | : :             | :|  | |  | |  :| :    | |     |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        473 PPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERA 532
              ||:     | |   :   | | | | | |  | |          ||::||::||: :: || |: :|
Db         95 PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        533 AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 592
                 |||:|: |||   :    : |: ||: ||  |   :| | | ||    |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        593 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 651
               :|| ::::|| ||::    |:||||    |  |:::||:||: |||||| |||||| | |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        652 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 703
              ||||::|||||||||::            |||| :  |||||  ||:      |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        704 EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761
              |   :| :        |      |||  |||||   | |:||||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTE 821
               | |||:|  :||||  |      | |:: |  |:|| |||||  :: |: :  || :  ||    || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        822 IVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 875
              |||||||||  :    |:  :  : ||||||| ||:               ||           |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        876 FGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLR 935
              ||||: ||||:|:: |    |  : |:|||: |    |  : |            || ::: :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        936 YYNEECRNCSTPGP------DIKLYPFIYHAVESCTQTTNQPGQRTGA 977
              |:   :  |:  |          ||| ||    |   :  : ||:   |
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGASA 606

Search completed: May 26, 2004, 18:39:56
Job time : 118.936 secs
```

```
OM protein - protein search, using sw model

Run on:          May 26, 2004, 18:46:21 ; Search time 81.5151 Seconds
                                           (without alignments)
                                           3343.018 Million cell updates/sec

Title:           US-09-302-812-2
Perfect score:   5184
Sequence:        1 MSAGPGCEPCTKRPRWDAAA..........YHAVESCTQTTNQPGQRTGA 977

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters:      1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                 8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                 10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                 12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                 13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                 14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                 15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                 16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                 17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                 18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5184	100.0	977	9	US-09-973-451-2	Sequence 2, Appli
2	4635.5	89.4	976	9	US-09-973-451-4	Sequence 4, Appli
3	4332.5	83.6	968	9	US-09-973-451-6	Sequence 6, Appli
4	1063.5	20.5	768	9	US-09-973-451-8	Sequence 8, Appli
5	594	11.5	546	12	US-10-425-114-60000	Sequence 60000, A
6	465	9.0	726	9	US-09-973-451-10	Sequence 10, Appl
7	205.5	4.0	200	12	US-10-424-599-184988	Sequence 184988,
8	184	3.5	180	12	US-10-424-599-156445	Sequence 156445,
9	167	3.2	31	9	US-09-973-451-11	Sequence 11, Appl
10	166	3.2	1462	14	US-10-287-218-17	Sequence 17, Appl
11	164.5	3.2	300	12	US-10-424-599-233915	Sequence 233915,
12	162.5	3.1	747	12	US-10-425-114-67803	Sequence 67803, A
13	162.5	3.1	1109	12	US-10-425-114-72939	Sequence 72939, A
14	162	3.1	748	9	US-09-864-761-43244	Sequence 43244, A
15	161	3.1	691	15	US-10-108-260A-4345	Sequence 4345, Ap
16	159	3.1	6642	15	US-10-369-493-5013	Sequence 5013, Ap
17	158.5	3.1	1973	16	US-10-197-824-34	Sequence 34, Appl
18	158.5	3.1	2047	12	US-10-346-863-2	Sequence 2, Appli
19	158.5	3.1	3863	16	US-10-197-824-7	Sequence 7, Appli
20	157.5	3.0	1163	12	US-10-221-278-197	Sequence 197, App
21	157.5	3.0	1163	15	US-10-291-172-197	Sequence 197, App
22	156	3.0	2665	9	US-09-864-761-34248	Sequence 34248, A
23	156	3.0	3664	12	US-10-263-929-143	Sequence 143, App
24	156	3.0	3664	14	US-10-177-293-423	Sequence 423, App
25	155.5	3.0	566	12	US-10-424-599-172054	Sequence 172054,
26	155.5	3.0	1501	9	US-09-924-154-17	Sequence 17, Appl
27	155.5	3.0	1804	14	US-10-275-140-2	Sequence 2, Appli
28	153	3.0	913	13	US-10-117-604-4	Sequence 4, Appli
29	152	2.9	914	15	US-10-369-493-1851	Sequence 1851, Ap
30	152	2.9	2697	15	US-10-144-198-12	Sequence 12, Appl
31	151.5	2.9	479	15	US-10-369-493-18078	Sequence 18078, A
32	151.5	2.9	665	10	US-09-820-843A-107	Sequence 107, App
33	149.5	2.9	1743	10	US-09-882-227-624	Sequence 624, App
34	149.5	2.9	1927	12	US-10-282-122A-58729	Sequence 58729, A
35	149.5	2.9	2224	14	US-10-115-563-14	Sequence 14, Appl
36	149	2.9	29	9	US-09-973-451-12	Sequence 12, Appl
37	148	2.9	3551	12	US-10-263-929-144	Sequence 144, App
38	147.5	2.8	1489	15	US-10-115-482-18	Sequence 18, Appl
39	147	2.8	26	9	US-09-973-451-19	Sequence 19, Appl
40	146.5	2.8	845	12	US-10-205-331-61	Sequence 61, Appl
41	146.5	2.8	1210	14	US-10-205-823-264	Sequence 264, App
42	146	2.8	1974	15	US-10-369-493-6395	Sequence 6395, Ap
43	145.5	2.8	447	10	US-09-794-422-44	Sequence 44, Appl
44	145.5	2.8	677	14	US-10-060-036-180	Sequence 180, App
45	145.5	2.8	677	14	US-10-060-036-4552	Sequence 4552, Ap

ALIGNMENTS

```
RESULT 1
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVI
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-2

  Query Match             100.0%;  Score 5184;  DB 9;  Length 977;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  977;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy    121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy    181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy    241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy    301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy    361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy    421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy    481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy    541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy    601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy    661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy    721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy    781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy    841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy    901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy    961 VESCTQTTNQPGQRTGA 977
          |||||||||||||||||
Db    961 VESCTQTTNQPGQRTGA 977
```

RESULT 2
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-973-451-4

```
  Query Match             89.4%;  Score 4635.5;  DB 9;  Length 976;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy      1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
          |:||||||||||  || || || |||||||||| ||||||| ||| |||||||||  |  
Db      1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy     61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
          |:||||||||||||||||||||||||||||| ||||| |||||||| |||||||||||||
Db     60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy    121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
          |||:||| |||||  |||  || |||| |||||||||||||| |:|:||||||  |||||
Db    120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy    181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
          |||||||:|:|   ||||||:|||:|||||||| ||||||||| |||  |||:||||| |
Db    180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy    241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
          :||| | || |||:| ||| :||||| |||||||| || |:| |||| ||||||||||||
Db    240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy    301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
          |||||||||||||||||||||||||||||||| ||:|||||||||| |:|| | ||| | 
Db    300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy    361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
          ||||:|||||||||||| ||||:||| ||: ||||||||||||||:|||||||| ||:||
Db    360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy    421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
          |||:|||| | ||||||||||||:||||||||||||||||||||||||||||| ||||||
Db    420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy    481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
          ||||||||||| ||||||||||:|||||||||||||||||||||||||||| ||||||||
Db    480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy    541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
          ||||||: ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy    601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy    661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy    721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
          |||||||| ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db    720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
```

```
Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
              ||||||||||||||||||||||:|||||  |||||  :|| ||||||||||||||||||||
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
              |:|||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
              ||||||||||||||| ||||||||||:|||||||||||||||||||||||||||||||||
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        961 VESCTQTTNQPGQRTG 976
              |||| :| :  |||||
Db        960 VESCAETADHSGQRTG 975
```

RESULT 3
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-973-451-6

```
  Query Match            83.6%;  Score 4332.5;  DB 9;  Length 968;
  Best Local Similarity  84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60
              ||||| |||||   || || || | |||:||||||||||| ||||||||||||    |
Db          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120
              |||| |||:||| |||||||:||||| || |||    ||||||| :|||||||||||
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180
               | :|||||| ||   |||| || :||| | |||::||||||| |:|:|| ||||| | :|
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
              :| ||||: ||   ||||||: || |||||||| :|||| | |: | |    |||: ||
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233

Qy        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
                :   : | |||||| ||: |||||| |:||:||| || |:|  |||||:||||||||
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360
              |||||||||:|||||||||||||| ||||:| || ||||| |   | |||| :|||
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352

Qy        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420
              || |:||||||| ||: || :|:|||   |::||||||||:|||||::||||||||||: |
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480
              :||:||||::||||||||||||||||:| |:|||||||||||||:||||| || |||||:
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540
              ||||:||||| ||||||||||:||||||||||||||||||||||||||||||| |||||||
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
              |||||: |||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
              ||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840
              ||||||||||||||||||||||||||| ||:||||| ||||||||||||||||||||||
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
              |:||||||||||||||| ||||||||||||||||||||||||||||||||| |||||||
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960
              ||||||||||||||||||||||||| ||:|||||||||||||||||||||||||||||||
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        961 VESCTQTTNQPGQRTG 976
              |||  :||: |||: |
Db        952 VESSAETTDMPGQKAG 967
```

RESULT 4
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-973-451-8

```
  Query Match             20.5%;  Score 1063.5;  DB 9;  Length 768;
  Best Local Similarity   41.5%;  Pred. No. 5e-74;
  Matches  244;  Conservative   93;  Mismatches  198;  Indels   53;  Gaps   15;

Qy        417 RVPKAEDKRKEQCEMKHQRTERKIPKYIPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 472
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        473 PPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERA 532
Db         95 PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        533 AGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 592
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        593 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 651
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        652 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 703
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        704 EDFPEWERCEKLL--TRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTE 821
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        822 IVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 875
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        876 FGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLR 935
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        936 YYNEECRNCSTPGP------DIKLYPFIYHAVESCTQTTNQPGQRTGA 977
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGASA 606

RESULT 5
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

  Query Match             11.5%;  Score 594;  DB 12;  Length 546;
  Best Local Similarity   33.8%;  Pred. No. 1.6e-37;
  Matches  153;  Conservative   72;  Mismatches  157;  Indels   70;  Gaps   12;

Qy        545 LNRLTRPQNLKDAILKYNVAYS------KKWDFTALIDFWDKVLEEAEAQHLYQSILPDM 598
Db         79 VSRVASGDVLADALTDLRLALALPALPPRAADGLAL--FFDDLLSRAQARGWFSEVVPNL 136

Qy        599 VKIALCLPNICTQPIPLLKQKMNHS--------------ITMSQEQIASLLANAFFCTFP 644
Db        137 ARLLLRLPTLLEDHY----AKAGHGASGLRVMASQDAGVVLLSQELVAALLTCALFCLFP 192

Qy        645 RRNAKMKSEYSSYPDINFNRLF-----EGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFT 699
Db        193 TAGRAQ----ACLPTINFDGLFAPLIHNARQSQE-QKVRCLVHYFERVTDSMPTGLVSFE 247

Qy        700 RQSL-----------EDFPEWERCEKLLTRLHVTYEGTIEGNGQGMLQVDFANRFVGGGV 748
Db        248 RKVLPRRALSDGVPYPDIHAWVASSAPLCQFTVFSSGFIEDEEQEALQVDFANKYLGGGA 307

Qy        749 TSAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHED 808
Db        308 LSWGCVQEEIRFMINPELIVGMLFLSCMEDNEAIEIFGAERFSQYMGYGSSFRFVGDYLD 367

Qy        809 RSERDDWQRRTTEIVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLR------------- 855
Db        368 TKPFDSMGRRRTRIVAIDALDCPARL-HYESGCLLREVNKAFCGFFDQSKHHLYAKLFQD 426

Qy        856 -------PGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVV-YFTFGDSE 907
Db        427 LHNKDDFSSINSSEYVGVSTGNWGCGAFGGNPEIKSMIQWIAASQALRPFVNYYTFEDVS 486

Qy        908 LMRDIYSMHTFLTERKLTVGEVYKLLLRYYNE 939
Db        487 LQR-LEEVIQWIRLHGWTVGELWHMLMEYSSQ 517

RESULT 6
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-973-451-10

  Query Match              9.0%;  Score 465;  DB 9;  Length 726;
  Best Local Similarity   24.3%;  Pred. No. 3.1e-27;
  Matches  204;  Conservative  116;  Mismatches  282;  Indels  236;  Gaps   35;

Qy        199 DTNSEESRDNQQFLTHVKLANAKQTME----DEQGREARSHQKCGKACHPAEACAGCQQE 254
Db         10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRRKLTEHGNTTESKE------DPEE------PK 56

Qy        255 ETDVVSESPLSDTGSED----------VGTGLKNA-------------NRLNRQESSLGN 291
```

```
Db         57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDN 116

Qy        292 ---SPPFEKESEPESPMDV------------DNSKNSCQDSEADEETSPGF--DEQE--- 331

Db        117 HKSTEPMEEDVNNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIE 176

Qy        332 -------------DSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESR 377

Db        177 YPGIMKDTTTQLDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSK 219

Qy        378 AGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTE 437

Db        220 A----TKTVRTSSSSFLST---------------------VSTCEAPAKGRARMYQKELE 254

Qy        438 RKIPKYIPPHLS--PDKKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRIGEV 495

Db        255 KHVIAFTEGNLTLQPD--------------------LNKVDPDRNY---------RYCTI 285

Qy        496 PKPFP-THFKDLWDNKH---VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRP 551

Db        286 PN-FPASQGKLREDNRYGPKIVLPQRWREF----DSRGRR-------------------- 320

Qy        552 QNLKDAILKYN---VAYSKKWDFTA-------LIDFWD---KVLEEAEAQHLYQSILPDM 598

Db        321 ---RDSYFYFKRKLDGYLKCYKTTGYFMFVGLLHNMWEFDPDITYKLPALEMYYKEMSEL 377

Qy        599 V-------------KIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPR 645

Db        378 VGREEVLEKFARVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF----- 431

Qy        646 RNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TRQSL 703

Db        432 --ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTKMDK 483

Qy        704 EDFPEWERCEKLLTRLHVTY--EGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761

Db        484 DTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEIRFL 542

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDRSE-RDDW 815

Db        543 MCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFRDRF 602

Qy        816 QRRTTEIVAIDALHFRR-----YLDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVATGN 870

Db        603 GRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIVTGW 661

Qy        871 WGCGAFGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGE 928

Db        662 WGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLGK 719
```

RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(200)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

```
  Query Match             4.0%;  Score 205.5;  DB 12;  Length 200;
  Best Local Similarity   29.4%;  Pred. No. 9.1e-08;
  Matches   52;  Conservative   26;  Mismatches   70;  Indels   29;  Gaps    4;

Qy        626 MSQEQIASLLANAFFCTFPRRNAKMKSEYSSYP---------DINFNRLFEGRSSRKPEK 676

Db         33 LTQETSAPPPSRTLFCLFP---------VSDRPGIHLPMINSDVSSGSLYEDYSQKPEST 83

Qy        677 LKTLFCYFRRVTEKKPTGLVTFTRQSLE--------DFPE---WERCEKLLTRLHVTYEG 725

Db         84 TGRIAHYGQRISSEMPKGIVSFERKVLPFKNDSIHISYPDANFWSTSAIPLCRFEVHSSG 143

Qy        726 TIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNECL 782

Db        144 LIEDQSSGAAEVDLANKYLGGGALGRGCVQEEIRFMVSPELEAGMLFLPAMADNEAI 200
```

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

```
  Query Match             3.5%;  Score 184;  DB 12;  Length 180;
  Best Local Similarity   29.4%;  Pred. No. 3.7e-06;
  Matches   48;  Conservative   24;  Mismatches   41;  Indels   50;  Gaps    4;

Qy        845 ELNKAYCGFL-----RP------------------------------------------ 856

Db          7 EINKAFCGFLYQCKYQPYQKILQENGCTSALFYAATSTSMETDEGEISNHKITNSQNDYH 66

Qy        857 GVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERD-VVYFTFGDSELMRDIYSM 915

Db         67 GMDQGNNIGVATGNWGCGAFGGDPEVKTIIQWLAASQALRPFIAYYTFG-LEALQSLDEV 125

Qy        916 HTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIY 958

Db        126 AHWILSQRWTVGDLWNMLIEYSINRSKGETNVGFLQWLLPSIY 168
```

RESULT 9
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,

```
; TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION:  THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE:  2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-11

  Query Match            3.2%;  Score 167;  DB 9;  Length 31;
  Best Local Similarity  100.0%;  Pred. No. 5.7e-06;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        771 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 801
              |||||||||||||||||||||||||||||||
Db          1 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 31
```

```
RESULT 10
US-10-287-218-17
; Sequence 17, Application US/10287218
; Publication No. US20030198975A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  AZIMZAI, Yalda; AU-YOUNG, Janice K.
;  APPLICANT:  BATRA, Sajeev; BAUGHN, Mariah R.
;  APPLICANT:  BECHA, Shanya D.; BOROWSKY, Mark L.
;  APPLICANT:  BUFORD, Neil; DING, Li
;  APPLICANT:  ELLIOTT, Vicki S.; EMERLING, Brooke M.
;  APPLICANT:  GANDHI, Ameena R.; GIETZEN, Kimberly J.
;  APPLICANT:  GRIFFIN, Jennifer A.; HAFALIA, April J.A.
;  APPLICANT:  HONCHELL, Cynthia D.; LAL, Preeti G.
;  APPLICANT:  LEE, Soo Yeun; LU, Dyung Aina M.
;  APPLICANT:  ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  REDDY, Roopa; SANJANWALA, Madhu, M.
;  APPLICANT:  TANG, Y. Tom; WALIA, Narinder K.
;  APPLICANT:  WANG, Yu-mei, E.; WARREN, Bridget A.
;  APPLICANT:  XU, Yuming; YANG, Junming
;  APPLICANT:  YAO, Monique G.; YUE, Henry
;  APPLICANT:  ZEBARJADIAN, Yeganeh
;  TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
;  FILE REFERENCE: PI-0417 USA
;  CURRENT APPLICATION NUMBER: US/10/287,218
;  CURRENT FILING DATE:  2002-10-31
;  PRIOR APPLICATION NUMBER: PCT/US02/11152
;  PRIOR FILING DATE: 2002-04-05
;  PRIOR APPLICATION NUMBER: US 60/349,705
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/295,263
;  PRIOR FILING DATE: 2001-06-01
;  PRIOR APPLICATION NUMBER: US 60/295,340
;  PRIOR FILING DATE: 2001-06-01
;  PRIOR APPLICATION NUMBER: US 60/293,727
;  PRIOR FILING DATE: 2001-05-25
;  PRIOR APPLICATION NUMBER: US 60/291,846
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/291,662
;  PRIOR FILING DATE: 2001-05-16
;  PRIOR APPLICATION NUMBER: US 60/287,228
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/286,820
;  PRIOR FILING DATE: 2001-04-26
;  PRIOR APPLICATION NUMBER: US 60/283,294
;  PRIOR FILING DATE: 2001-04-11
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 17
;   LENGTH: 1462
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030198975A1 5734806CD1
US-10-287-218-17

  Query Match            3.2%;  Score 166;  DB 14;  Length 1462;
  Best Local Similarity  19.7%;  Pred. No. 0.0022;
  Matches  136;  Conservative   97;  Mismatches  253;  Indels  204;  Gaps   28;

Qy         39 VLDSKDAPVQFRVPPSSSGC--ALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESES 96
Db        606 VIPLSDSPPPSRWKPGQKPWKPSYERIQEMKAKTTHLLPIQSTY-SLANIK--ETGSSSS 662

Qy         97 LHSKENNNTREESMMSSV---------------QKDNFYQHNMEKLENV-SQLGFDKSPV 140
Db        663 YHKREKNSESDQSTYSKYSDRSSESSPRSRSRSSRSRSYSRSYTRSRSLASSHSRSRSPS 722

Qy        141 EKG---TQYLKQHQTAAMCKW----QNEGPHSERLLESEPPAVTLVPEQFSNANV----- 188
Db        723 SRSHSRNKYSDHSQCSRSSSYTSISSDDGRRAKRRLRSSGKKNSVSHKKHSSSSEKTLHS 782

Qy        189 ------DQSSPKDDHSDTNS--EESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        783 KYVKGRDRSSCVRKYSESRSSLDYSSDSEQ--SSVQATQSAQEKE-KQGQMERTHNKQEK 839

Qy        241 ACHPAEACAGCQQEETDVVSESPLSD--TGSEDVGTGLKNANRLNRQESSLGNSPPFEKE 298
Db        840 ---------NRGEEKSKSERECPHSKKRTLKENLSDHLRNGSKPKRK-----NYAGSKWD 885

Qy        299 SEPESPMDV-DNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKR 357
Db        886 SESNSERDVTKNSKNDSHPSSDKEEGEATSDSESEVSEIHIKVKPT-----------TK 933

Qy        358 SSAKGGEIRLHFQFEGGESRAGMNDVN-----AKRPGSTSSLNVECRNSK---------Q 403
Db        934 SS---------------TNTSLPDDNGAWKSSKQRTSTSDSEGSCSNSENNRGKPQKHK 977

Qy        404 HGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMR 463
Db        978 HGSKENLKREH---TKKVKEKLKGKKDKKHKAPKRKQAFHWQP---------------- 1017

Qy        464 RMPRCGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYP 523
Db       1018 --------------------------------PLEFGEEEEEEIDDKQVTQESKEKKV-- 1043

Qy        524 VEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVA----YSKKWDFTALIDFWD 579
Db       1044 --SENNE----------------------TIKDNILKTEKSSEEDLSGKHDTVTVSSDLD 1079

Qy        580 KVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAF 639
Db       1080 QFTKDDSKLSISPTALNTEENVA-CLQNI---------QHVEESVPNGVEDVLQTDDNME 1129

Qy        640 FCTFPRRNAKMKSEYSS--------YPDIN 661
Db       1130 ICT-PDRSSPAKVEETSPLGNARLDTPDIN 1158
```

```
RESULT 11
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
```

```
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233915
;   LENGTH: 300
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_5324C.1.pep
US-10-424-599-233915

  Query Match             3.2%;  Score 164.5;  DB 12;  Length 300;
  Best Local Similarity   18.0%;  Pred. No. 0.00027;
  Matches   53;  Conservative   47;  Mismatches   66;  Indels  129;  Gaps    5;

Qy        577 FWDKVLEEAEAQHLYQSILPDMVKIALCLPNI--------------------CTQPIPLL 616
Db         76 FFDEVMSGEESSKWFQEVLPALGNLLLRLPSLLESHYQNTDNMAIDGEAGAMLTTALRLL 135

Qy        617 KQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYP---------DINFNRLFE 667
Db        136 DSQQPGIVFLTQELIAALLSCSLFCLFP---------VSDRPVIHLPMINFDVLFGSLYD 186

Qy        668 GRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLHVTYEGTI 727
Db        187 DYSQKQENKIWCIVHYFQRISSEMPKGIA------------------------------ 215

Qy        728 EGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGT 787
Db        216 --------------------------------------------------YNERIEIVGV 225

Qy        788 EQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPEKI 842
Db        226 ERFSGYTDHASSFRFSRAKAEGREEDPVGRRKTSDLS----------EKYFPPKM 270

RESULT 12
US-10-425-114-67803
; Sequence 67803, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67803
;   LENGTH: 747
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-225-C11_FLI.pep
US-10-425-114-67803

  Query Match             3.1%;  Score 162.5;  DB 12;  Length 747;
  Best Local Similarity   20.4%;  Pred. No. 0.0015;
  Matches  101;  Conservative   72;  Mismatches  221;  Indels  101;  Gaps   15;

Qy         19 AATSPPAASDA------------RSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRAGQH 66
Db         17 AAVSVPEGKSAKKGKRNAEDEIEKAVSAKKQKTVPEKAVPVKKQPPPKKVESSSSEEDSS 76

Qy         67 RGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESM--MSSVQKDNFYQHNM 124
Db         77 DSEVEVKVQPKKVIQPKKSTQPAKQ-ESSDDSSSESSSDDEPSKKPVAGSKKPPASSSSS 135

Qy        125 EKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFS 184
Db        136 DDSSDSDDSSSDEEPAKKPTVLSKK---------------PTVLSKKPTVLSKKPVTAV 179

Qy        185 NANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHP 244
Db        180 SNGSKQVKPDSSSSDSSSDEDEKPAASLKKPVVASVQKKTQESDSSDSDSDEDESEDDIP 239

Qy        245 AEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNR----------QESSLGNSPP 294
Db        240 AKAQVVAKKKEESSSSSDSDSDSESEDE----KPAANLKKPPVPSVQKKTQESDSSDSDS 295

Qy        295 FEKESE--PESPMDVDNSK-NSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREAD 351
Db        296 DEDESDDIPAKAQVVAKKKEESGSSSDSDSE----FESDDEDSNAKTV------QPAKAA 345

Qy        352 TELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVE------------- 397
Db        346 AIKKKEDSSESSE--TDSQSDSEDSEPEKPTVPAKRPLATNKKNEQSSEKSDDSSDESSD 403

Qy        398 ----------CRNSKQHG----------RKDSKITDHFMRVPKAEDKRKEQCEMKHQR-- 435
Db        404 ESDEEPPQKKTKNSAPSGAAKAATNVSKKKSSSDEDDESSEESSDDEDTEQVETKETSVV 463

Qy        436 TERKIPKYIPPHLSP 450
Db        464 TTQK-----PAHNEP 473

RESULT 13
US-10-425-114-72939
; Sequence 72939, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72939
;   LENGTH: 1109
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4119-145-F11_FLI.pep
US-10-425-114-72939

  Query Match             3.1%;  Score 162.5;  DB 12;  Length 1109;
  Best Local Similarity   18.8%;  Pred. No. 0.0027;
  Matches  138;  Conservative  111;  Mismatches  308;  Indels  177;  Gaps   29;

Qy         76 KQKTITSWMDTKGIKTVESESL-HSKENN------NTREESMMSSVQKDNFYQHNMEK-- 126
Db        438 KRPKLENLPDTEDQETVDINSVSEGKENNIMITLETNIEHNLKSEEEKDQEKQQMFENKL 497

Qy        127 ----------LENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAV 176
Db        498 IKSEEIKDTILQTVDLVSQETGEKEANIQAVDSE--VGLTKEDTQ----EKLGEDDKTQK 551

Qy        177 TLVPE------QFSNANVDQSSPKDDHSDTNSEESRD----NQQFLTH---------VKL 217
Db        552 DVISNTSDVIGTYEAADVAQKVDEDSAEDTQSNDGKEVVEVGQKLINKPMVGPEAGGTKE 611
```

```
Qy        218 ANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQE----ETDVVSESP---------- 263
                  |: :|  :  | ::  :    :| : :|       :|    | : ::||
Db        612 VPIKEIVEMNEIEEGKNKE---QAINSSENIMDINEEPGTTEGEEITESSSTEEMEVRSV 668

Qy        264 LSDTGSEDVGTGLKNANRLNRQ--------ESSLGNSPPFEKESEPESPMDV-DNSKNSC 314
              ::||  : :|: :::|:::  |           |:   |     |:|  |   |   |  |
Db        669 VADTDQKALGSEVQDASKVTTQIDKEKKEIPVSIKKEPEVTVVSQPTEPQPVLIPSININ 728

Qy        315 QDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEG- 373
               ||  ::|       |     |   :|    |     |:| | :     |:|    | |:
Db        729 SDSGENKE------EIGSLSKTETILPPESENPKENDNDSGTGSTADTSSIDLNLSISSF 782

Qy        374 ------------GESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITD-HFMR--- 417
                          |:|      :      |         :  |   ||     |||  :  |:|
Db        783 LSKTKDSGSISLQETRRQKKTLKKTRKFIVDGVEVSVTTSKIVTDSDSKTEELRFLRRQE 842

Qy        418 ------VPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIR 471
                     : |  |  :  ::|   |  |:   :|  :    :   |:   ||  : : :   |
Db        843 LRELRFLQKEEQRAQQQLNSKLQQQREQIFRRFEQEMMSKKRQYDQEIENLEKQQKQTIE 902

Qy        472 LPPLRPSANHTVTIRVDLLRI-GEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGE 530
                     |    ||  :|  :   || ||   |    : |:::  |:  :    ||     :  :  :
Db        903 ----RLEQEHTNRLRDEAKRIKGEQEKEL-SKFQNMLKNRKKE----EQEFVQKQQQELD 953

Qy        531 RAAGSRWELIQ---TALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEA 587
                  ||   ::||     | |        | |   :::    |                 |:   |||
Db        954 ---GSLKKIIQQQKAELANIERECLNNKQQLMRAREA-----------AIWE--LEERHL 997

Qy        588 QHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRN 647
              |   :|                        ||||::      |  : |:              :|:
Db        998 QEKHQ---------------------LLKQQLKDQYFMQRHQLL------------KRH 1023

Qy        648 AKMKSEYSSYPDINFNRLFEGRSSRKPE-----------KLKTLFCYFRRVTEKKPTGLV 696
               |      :      |         ||  |     :|: :                 :  ||      |::           |
Db       1024 EKETEQMQRYN----QRLIEELKNRQTQERARLPKIQRSEAKTRMAMFKKSLRINSTATP 1079

Qy        697 TFTRQSLEDFPEWE 710
                  |   :: |   |
Db       1080 DQDRDKIKQFAAQE 1093
```

RESULT 14
US-09-864-761-43244
; Sequence 43244, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43244
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004886.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: AI636014.1, EVALUE 3.00e-51
; OTHER INFORMATION: SWISSPROT HIT: P38254, EVALUE 1.40e+00
US-09-864-761-43244

```
  Query Match             3.1%;  Score 162;  DB 9;  Length 748;
  Best Local Similarity   18.4%;  Pred. No. 0.0017;
  Matches  115;  Conservative  109;  Mismatches  225;  Indels  176;  Gaps   26;

Qy        105 TREESMMSSVQKDNFYQH-------NMEKLEN---------VSQLGFDKSPVEKGTQYLK 148
               | |: ::  ::||:|  |               :: |||:                 |  |   :||   || ||
Db         44 TDEKDILKGLKKDSFSQESSPSSPSDLAKLESTVLSILEAQASTLADEKS--EKKTQ--- 98

Qy        149 QHQTAAMCKWQNEGPHSERLLE--SEPPAVTLVPEQFSNANVDQSS--PKDDHSD-TNSE 203
               |: :          | |   :    :  ||    :|:  |:    :   ::         ||    |   :|:
Db         99 PHEVSP------EQPKDQEKTQSLSETLEITISEEEIKESQEERKDTFKKDSQQDIPSSK 152

Qy        204 ESRDNQQFLTHVKLA-----NAKQTMEDEQG----REARSHQKCGKACHPAEACAGCQQE 254
              : ::  :|:   :           : :::  |  |          |:  |:            :       :    :     |
Db        153 DHKEKSEFVDDITTRREPYDSVEESSESENSPVPQRKRRTSVGSSSSDEYKQEDSQGSGE 212

Qy        255 ETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPE----------- 302
               | |  :   :   :      ||    :|  ::     :       |     : :|     :|:  | :
Db        213 EEDFIRKQIIEMSADED-ASGSEDDEFIRNQLKEISSSTESQKKEETKGKGKITAGKHRR 271

Qy        303 ------SPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPS------------R 344
                     : :|  |   :          | ||      |||  :     :|  :: |                    |
Db        272 LTRKSSTSIDEDAGRRHSWHDEDDE----AFDESPELKYRETKSQESEELVVTGGGGLRR 327

Qy        345 FQPREADTELRKRSSAKGGEIRLHFQF---------------------EGGESRAGMNDV 383
               |:  | :: :   : ||:   :  :    |                                       ||   |      :
Db        328 FKTIELNSTIADKYSAESSQKKTSLYFDEEPELEMESLTDSPEDRSRGEGSSSLHASSFT 387

Qy        384 NAKRPGSTSSLNVECRNSKQHGRKDSK------ITDHFMRVPKAEDK------RKEQCEM 431
                   |  |  |||:  :   :|   |  : :||              |     :|   ||             |:|:   :
Db        388 PGTSPTSVSSLDEDSDSSPSHKKGESKQQRKARHRPHGPLLPTIEDSSEEEELREEEELL 447

Qy        432 KHQRTERKI--------------------------PKYIPPHLSPDKKWLGTPIEEMRR 464
               |  |    :|:|                             ||    | :|||  :        :|  ||:|:
Db        448 KEQEKQREIEQQQRKSSSKKSKKDKDELRAQRRRERPKTPPSNLSPIED--ASPTEELRQ 505

Qy        465 MPRCGIRLPPLRPS--ANHTVTIRVDLLRIGEVPKPFPTHFKDLWD-NKHVKMPCSEQNL 521
                           :    |    |    :  :: :|     |              |:  |       :  :  :      |    |:|   :
Db        506 ----AAEMEELHRSSCSEYSPSIESD-------PEGFEISPEKIIEVQKVYKLPTAVSLY 554

Qy        522 YPVEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTAL--IDFWD 579
                 |  ::::   :    ||:                              :  ||  |       |        |      :  |        :     |
```

```
Db        555 SPTDEQSIMQKEGSQ---------------KALKSAEEMYEEMMHKTHKYKAFPAANERD 599

Qy        580 KVLEEAE-------AQHLYQSILPD 597
               :| |:          ::|:|:: |
Db        600 EVFEKEPLYGGMLIEDYIYESLVED 624


RESULT 15
US-10-108-260A-4345
; Sequence 4345, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4345
;   LENGTH: 691
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-4345

  Query Match              3.1%;  Score 161;  DB 15;  Length 691;
  Best Local Similarity   19.6%;  Pred. No. 0.0018;
  Matches  143;  Conservative  101;  Mismatches  291;  Indels  196;  Gaps   35;

Qy         41 DSKDAPVQFRVPPSSSGCAL---GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESES- 96
Db         31 DTSGSSVARRAPKRQASCILNVQSRSGDTSGSSVARRAPKRQASS------VVVIDSDSD 84

Qy         97 --LHSKENN-------NTREES---------------MMSSVQKDNFYQHNMEKLENVSQ 132
Db         85 EECHTHEEKKAKLLEINSDDESPECCHVKPAIQEPPIVISDDDNDDDNGNDLEVPDDNSD 144

Qy        133 LGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSS 192
Db        145 --------------------------DSEAP-DDNSDDSEAPDDNSDDSEAPDDNSDDSE 177

Qy        193 PKDDHS------DTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAE 246
Db        178 APDDNSDDSDVPDDNSDDSSDD----------NSDDSSDDNSDDSDVPDDKSDDSDVPDD 227

Qy        247 ACAGCQQEETDVV------SESP---LSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEK 297
Db        228 S-----SDDSDVPDDSSDDSEAPDDSSDDSEAPDDSSDDSEAPDDSSDDSEAPDDSSDDS 282

Qy        298 ESEPESPMDVDNSKNSCQDSEADEETSPGFD---EQEDSSSAQTANKPSRFQPREADTEL 354
Db        283 EASDDSSDDSEASDDSSDDSEAPDDKSDDSDVPEDKSDDSDVPDDNSDDLEVPVPAEDLC 342

Qy        355 RKRSSAKGGE--------IRLHFQFEGGESRAGMNDV--NAKRPGSTSSLNVECRNSK-- 402
Db        343 NEGQIASDEEELVEAAAAVSQH----DSSDDAGEQDLGENLSKPPSDPEANPEVSERKLP 398

Qy        403 ----------QHGRKDSKITDHFMRVPK---------AEDKRKEQCEMKH-------QRT 436
Db        399 TEEEPAPVVEQSGKRKSK-TKTIVEPPRKRQTKTKNIVEPPRKRQTKTKNIVEPLRKRKA 457

Qy        437 ERKIPKYIPPH--LSPDKKWLGTPIEEMR--RMPRC---GIRLPPLRPSANHT------- 482
Db        458 KTKNVSVTPGHKKRGPSKKKPGAAKVEKRKTRTPKCKVPGCFLQDLEKSKKYSGKNLKRN 517

Qy        483 ----VTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMP--CSEQNL-YPVEDENGERAAGS 535
Db        518 KDELVQRIYDLFNRSVCDKKLPEKLRIGWNNKMVKTAGLCSTGEMWYP---------KWR 568

Qy        536 RWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALI----DFWDKVLEEAEAQHLY 591
Db        569 RFAKIQIG-LKVCDSADRIRDTLI-HEMCHAASWLIDGIHDSHGDAWKYYARKSNRIH-- 624

Qy        592 QSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIA----SLLANAFFCTFPRRN 647
Db        625 ----PELPRVTRC-HNY------KINYKVHYECTGCKTRIGCYTKSLDTSRFIC------ 667

Qy        648 AKMKSEYSSYP 658
Db        668 AKCKGSLVMVP 678

Search completed: May 26, 2004, 19:18:52
Job time : 84.5151 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.8764 Seconds (without alignments) 3496.721 Million cell updates/sec

Title: US-09-302-812-2
Perfect score: 5184
Sequence: 1 MSAGPGCEPCTKRPRWDAAA..........YHAVESCTQTTNQPGQRTGA 977

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	522	10.1	997	2	B84726	probable poly(ADP-
2	465	9.0	726	2	T21138	hypothetical prote
3	331	6.4	364	2	A84726	probable poly(ADP-
4	185.5	3.6	1320	2	S57113	BUD4 protein - yea
5	175	3.4	1359	2	T34036	hypothetical prote
6	166	3.2	1403	1	A47328	natural killer cel
7	164	3.2	1115	2	T19137	hypothetical prote
8	162.5	3.1	1071	2	E85343	hypothetical prote
9	161.5	3.1	646	1	S15901	chromogranin B pre
10	159	3.1	1535	2	T49042	hypothetical prote
11	159	3.1	6642	2	T29757	protein UNC-89 - C
12	158.5	3.1	1576	2	T03277	pol protein - yeas
13	158	3.0	489	2	A45988	dentin matrix acid
14	157.5	3.0	1231	2	T18532	serine/threoine pr
15	157	3.0	1192	2	A71623	probable secreted
16	156.5	3.0	1200	2	A46194	neurofilament prot
17	155.5	3.0	522	2	C96608	hypothetical prote
18	155.5	3.0	1463	2	T30290	AAS surface protei
19	155.5	3.0	1804	2	T34518	nestin - golden ha
20	154.5	3.0	611	2	T06458	nucleolin homolog
21	154	3.0	406	2	S38170	SRP40 protein - ye
22	153	3.0	3122	2	T17202	DNA-directed DNA p
23	153	3.0	5327	2	T13564	microtubule-associ
24	152.5	2.9	915	2	B59433	chromosome 5 GAP-1
25	152	2.9	914	2	S48333	ORC1 protein - yea
26	152	2.9	1507	2	B47328	natural killer cel
27	151.5	2.9	665	2	B71609	hypothetical prote
28	151.5	2.9	1213	2	A58198	serine/proline-ric
29	151	2.9	1110	2	I51116	NF-180 - sea lampr
30	151	2.9	2248	2	A35938	profilaggrin - hum
31	150.5	2.9	532	2	T06029	hypothetical prote
32	150.5	2.9	1269	2	F84730	probable myosin he
33	150.5	2.9	1298	2	I54367	X-linked nuclear p
34	149.5	2.9	1927	2	G64585	cag pathogenicity
35	149.5	2.9	3147	2	T18674	hypothetical prote
36	148	2.9	635	2	T09648	nucleolin homolog
37	148	2.9	677	1	S09078	chromogranin B pre
38	148	2.9	3225	2	I52300	giantin - human
39	148	2.9	3259	1	A56539	giantin - human
40	147.5	2.8	913	2	T52485	neurofilament prot
41	146.5	2.8	845	2	A45669	neurofilament trip
42	146.5	2.8	1210	2	I39410	AF-4 protein, spli
43	146.5	2.8	1489	2	S60416	DNA helicase YGL15
44	146.5	2.8	1641	2	I38614	helicase II - huma
45	146.5	2.8	1791	2	T02345	hypothetical prote

ALIGNMENTS

RESULT 1
B84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31870
A;Map position: 2

```
  Query Match             10.1%;  Score 522;  DB 2;  Length 997;
  Best Local Similarity   32.0%;  Pred. No. 6.3e-24;
  Matches  136;  Conservative   59;  Mismatches  116;  Indels  114;  Gaps   12;

Qy        578 WDKVLEEAEAQHLYQSILPDMVKIALCLPNI--------------CTQPIPLLKQKMNHS 623
              :|::::| |:: :  |:| : : |  |::                 : ||  :
Db         79 FDELIDEKESKRWFDEIIPALASLLLQFPSLLEVHFQNADNIVSGIKTGLRLLNSQQAGI 138

Qy        624 ITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNR--LFEGRSSRKPEKLKTLF 681
              : :||| | :||| :||| || |  |       | |||:  |:  |  : |:: :
Db        139 VFLSQELIGALLACSFFCLFPDDNRGAK----HLPVINFDHLSLYISYSQSQESKIRCIM 194

Qy        682 CYFRRVTEKKPTGLVTFTRQ--SLEDFPEWERCEKLLTRLHVTYEGTIEGNGQGMLQVDF 739
               || |    | |:|:| |:  |   |  : :  |     |:             |:|||
Db        195 HYFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSL------YQ------PDNALEVDF 242

Qy        740 ANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAET 799
              ||:::|||  |  | ||||||:||||| ||   :| || : |  | |:: | ||||| :
Db        243 ANKYLGGGSLSRGCVQEEIRFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCYTGYASS 302

Qy        800 YRWARSHEDRSERDDWQRRTTEIVAIDAL------HFRRYLDQFVPEKIRRELNKAYCGF 853
              :|:|  : |:  | ::|| | ||||||     ||:           : ||:||| |||
Db        303 FRFAGEYIDKKAMDPFKRRRTRIVAIDALCTPKMRHFKDIC-------LLREINKALCGF 355

Qy        854 LR---------------------------------------------------------- 855
              |
Db        356 LNCSKAWEHQNIFMDEGDNEIQLVRNGRDSGLLRTETTASHRTPLNDVEMNREKPANNLI 415

Qy        856 -----PGVSSENL--SAVATGNWGCGAFGGDARLKALIQILAAAVAERDVV-YFTFGDSE 907
                   || :|:    |||||||| ||||  ||| || |||:   |  : |:|||  |
Db        416 RDFYVEGVDNEDHEDDGVATGNWGCGVFGGDPELKATIQWLAASQTRRPFISYYTFG-VE 474
```

```
Qy        908 LMRDI 912
                 :|::
Db        475 ALRNL 479

RESULT 2
T21138
hypothetical protein F20C5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T21138
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19381
A;Accession: T21138
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-726 <WIL>
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1
A;Experimental source: clone F20C5
C;Genetics:
A;Gene: CESP:F20C5.1
A;Map position: 4
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

  Query Match             9.0%;  Score 465;  DB 2;  Length 726;
  Best Local Similarity   24.3%;  Pred. No. 1.2e-20;
  Matches  204;  Conservative  116;  Mismatches  282;  Indels  236;  Gaps   35;

Qy        199 DTNSEESRDNQQFLTHVKLANAKQTME----DEQGREARSHQKCGKACHPAEACAGCQQE 254
               |  ::: :|  : ::  |   |:    ||:        ||     | :         ||        :
Db         10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRRKLTEHGNTTESKE------DPEE------PK 56

Qy        255 ETDVVSESPLSDTGSED----------VGTGLKNA-------------NRLNRQESSLGN 291
                 ||    |  ||   ||                 :|                 |  |  | |
Db         57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDN 116

Qy        292 ---SPPFEKESEPESPMDV------------DNSKNSCQDSEADEETSPGF--DEQE--- 331
                 : |  |::    :|  :||            :  :       :|  |    ::  |    |:||
Db        117 HKSTEPMEEDVNNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIE 176

Qy        332 --------------DSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESR 377
                                 |   :||  |     :   |||:                  |  | :|:
Db        177 YPGIMKDTTTQLDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSK 219

Qy        378 AGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTE 437
                          |   |:|  |:                        |    |    |  :  |   :  |
Db        220 A----TKTVRTSSSSFLST---------------------VSTCEAPAKGRARMYQKELE 254

Qy        438 RKIPKYIPPHLS--PDKKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRIGEV 495
               :  :    :  :|:  ||                      |   |  |:             |      :
Db        255 KHVIAFTEGNLTLQPD--------------------LNKVDPDRNY---------RYCTI 285

Qy        496 PKPFP-THFKDLWDNKH---VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRP 551
               |  ||  :    |      ||::    :  |      :           |  |
Db        286 PN-FPASQGKLREDNRYGPKIVLPQRWREF----DSRGRR-------------------- 320

Qy        552 QNLKDAILKYN---VAYSKKWDFTA-------LIDFWD---KVLEEAEAQHLYQSILPDM 598
                   :|:    :        |  |  :   |             |  :  |:     :     :   |    :|    : ::
Db        321 ---RDSYFYFKRKLDGYLKCYKTTGYFMFVGLLHNMWEFDPDITYKLPALEMYYKEMSEL 377

Qy        599 V-------------KIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPR 645
               |                   :||      :|   :  |   |      :  |  |:|  :|  |:|:|      ||
Db        378 VGREEVLEKFARVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFF----- 431

Qy        646 RNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TRQSL 703
                    |:    |  :|                  |     |||  ||  ||    :::          |  |  :|    |:
Db        432 --ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTKMDK 483

Qy        704 EDFPEWERCEKLLTRLHVTY--EGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 761
                :  |   |    :  :||  :       |   :     |    ||               ||||||     :||||  :   |  ||||||||
Db        484 DTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEIRFL 542

Qy        762 INPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDRSE-RDDW 815
               :  ||::|    |    |   :      |  :  |   |        :|   ||||    |  :||              |  ::  :|  ||  :
Db        543 MCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFRDRF 602

Qy        816 QRRTTEIVAIDALHFRR-----YLDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVATGN 870
                |      |  :||||:  |:                 :|          |  ||:  ||     ||:     |          |:     :   ||
Db        603 GRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIVTGW 661

Qy        871 WGCGAFGGDARLKALIQILAAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVGE 928
               ||||||  ||    ||  :||::||  ||:|  :  :  :||:  ||              :        :  ::  :|:|:
Db        662 WGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLGK 719

RESULT 3
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

  Query Match             6.4%;  Score 331;  DB 2;  Length 364;
  Best Local Similarity   25.3%;  Pred. No. 7.2e-13;
  Matches  110;  Conservative   63;  Mismatches  140;  Indels  122;  Gaps   12;

Qy        582 LEEAEAQHLYQSILPDMVKIALCLPNICTQ--------------PIPLLKQKMNHSITMS 627
               : : |:    :        ||  |   :  |    |::                              : :|          : :|
Db          1 MSKEESSRWFNEFLPAMACLLLRFPSLLESHYLNSDNLINGTKTGLRVLVPNKAGIVFLS 60

Qy        628 QEQIASLLANAFFCTFPRRNAKMKSEYSSY-PDINFNRLF-----EGRSSRKPEKLKTLF 681
               || |  :||: :|||  ||       :        |::  |  |||::||        ||:   :   |:|  :
Db         61 QELIGALLSCSFFCLFP-----VDDRGSNHLPIINFDKLFGSLINTGRNEHQENKIKCII 115

Qy        682 CYFRRVTEKKPTGLVTFTRQ--SLE------DFPEWERCEKLLTRLHVTYEGTIEGNGQG 733
                 ||:|::         |  :|  |:   |||              |           |            |     |       |   ||
Db        116 HYFQRLSSSISPGFVSFERKILSLEQDSSTLDEGFWGKSTVNLCPVEVRTSGLIEDQSVE 175

Qy        734 MLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEY 793
                |:|||||: :|||        |  |||||||:||||||    ||       ::     |   :  :   |  |::|   |
Db        176 ALEVDFANKNLGGGALRKGCVQEEIRFMINPELIVGMLFLPTMEVTEAIEVVGAERFSLY 235

Qy        794 TGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPEKIRRELNKAYCGF 853
               ||                                 :::    |    |||||||
Db        236 TGC-----------------FRKAKTRIVAIDALR------------------------ 253

Qy        854 LRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVYFT-------FGDS 906
                 ||||                                       :|::|:  :|     :    |    :               ||
Db        254 -HPGVS-------------------QYKLESLLSVLILSSSGRPIRLYMGSVSLQGIGDV 293

Qy        907 ELMRDIYSMHTFLTERKLTVGEVYKL---------------LLRYYNEECRNCSTPGPDI 951
                  ||  :|  |        |         :           :|                   |       ::   :     |      |  :
Db        294 VLMVEILSSSLFFNGLRFHRSNLYLFSPLIACNYQPMVVVWLTTLHSFQARR-----PFM 348

Qy        952 KLYPFIYHAVESCTQ 966
                 |  |  :  |:::   |
Db        349 SYYTFGFEALQNLNQ 363
```

```
RESULT 4
S57113
BUD4 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J1905; protein YJR092w
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 29-Oct-1999
C;Accession: S57113; S61952
R;Ramezani Rad, M.; Kirchrath, L.; Hollenberg, C.P.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57111
A;Accession: S57113
A;Molecule type: DNA
A;Residues: 1-1320 <RAM>
A;Cross-references: EMBL:Z49592; MIPS:YJR092w
R;Sanders, S.L.; Herskowitz, I.
submitted to the EMBL Data Library, November 1995
A;Reference number: S61952
A;Accession: S61952
A;Molecule type: DNA
A;Residues: 'MHDAESTVDSLLKEIDNEMEQTKSNITQNGSEDTPHNWKLPLQEIGDDTMEMLVKHNTRSNATENSRGRSP','S
202-212,'D',214-1007,'A',1009-1320 <SAN>
A;Cross-references: EMBL:U41641; NID:g1151116; PIDN:AAB17116.1; PID:g1151117
C;Genetics:
A;Gene: SGD:BUD4
A;Cross-references: MIPS:YJR092w; SGD:S0003852
A;Map position: 10R

  Query Match             3.6%;  Score 185.5;  DB 2;  Length 1320;
  Best Local Similarity   16.9%;  Pred. No. 0.0031;
  Matches  185;  Conservative  156;  Mismatches  361;  Indels  391;  Gaps   45;

Qy         70 ATSLVFKQKTITSWMDTK----GIKTVESESLHSKEN-----NNTREESMMSSVQKDNFY 120
Db        290 STSTKSEESYIADYKVTRQEDWDTKKLHQESEHANEQPAIIPQKDSSEETFTELNNESEF 349

Qy        121 QHNM-----------------------EKLENVSQLGFD-------KSPVEKGTQYLKQ 149
Db        350 QRNFKDGEEYRIVQHEESLYGQRTKSPEENIINGSEIGVDHGEAAEVNEPLAKTS--AEE 407

Qy        150 HQTAAMCKWQ--NEGPHSERLLE------------------------SEPPAVTLVPE- 181
Db        408 HDLSSSCEDQSVSEARNKDRIEEKEVETKDENIETEKDESEYHKVEENEEPEHVPLLPPL 467

Qy        182 ------QFSNANVDQSSPKDDHSD---------------------TNSEESRDNQQFLT 213
Db        468 PRWEEIQFNEPFIDENDTSNDSIDLTRSMKPSDYISIWHIQEEEIKSNSPESIANSQFSQ 527

Qy        214 HVKLANAKQTMEDEQGREARSHQ------KCGKACHPAEACAGCQ--QEETDVVSESP-- 263
Db        528 QSSITTA-STVDSKKDNGSTSFKFKPRIVSRSRIYNPKSRVSSLNYYDNEDYILSNSEWN 586

Qy        264 ---------LSDTGSEDVGTGLKNA-------NRLNRQESSLGNSPPFEKESEPESPMD 306
Db        587 ALDPMRRNTLISKRIQDNIRTQKGHAPLIRPSIMKLNGEDSGFQN-----HFLEVEQPQE 641

Qy        307 VDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA------ 360
Db        642 HENIPLSTHLSEQDITTNVGLDEQKLPTNTQD----------EAEISIREIESAGDITFN 691

Qy        361 KGGEIRLHFQFEGGESRAGMNDV----------------NAKRPGSTSSLNVECRNSKQ- 403
Db        692 RGDLLSLSFDEELGQDFANFLDALDHDSTSFNHGPDDSSSFQRDSSKKSFNSLWESSYEL 751

Qy        404 ----HGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTER-------------KIP----- 441
Db        752 KPPPSIRKQPIAPDVLQKL--LESDTKDDADLEKIREERITEPRTGLGIGMLKTPVKDVS 809

Qy        442 ------------------------------------------KYIPP----HLSPDKKWLG 456
Db        810 IALAASIKGYEASFSDTDSRPEGMNNSDAITLNMFDDFEEDKMTPSTPVRSISPIKRHVS 869

Qy        457 TP---------------------IEEMRRMPRCGIR--LPPLRPSANHTVTIRVDLLRI 492
Db        870 SPFKVVKAGNKQENNEINIKAEEEIEPMTQQETDGLKQDIPPLLAQTKDNVEAKEETITQ 929

Qy        493 GEVPKPFPTHFKD---------------LWDNKHVKMPCSEQNLYPVEDENGERAAGSRW 537
Db        930 LEEPQDVEQEFPDMGTLYLSIKAISTLALYGTK------SHRATYAIVFDNGENVVQTPW 983

Qy        538 ELI--------------------QTALLNRLTRPQNLKDAILKYNVAYSK-KWDFTALID 576
Db        984 ESLPYDGNIRINKEFELPIDFKGKTETSSASSERDSYKKCVITLKCKYEKPRHELVEIVD 1043

Qy        577 ---------------FWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMN 621
Db       1044 KVPVGKSFFGKTKYKFEKKYVQKKPKQDEWDYLFAQDGSFARC----------------- 1086

Qy        622 HSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPD--INFNRLFEGRSSRKPEKLKT 679
Db       1087 -EIEINEE----FLKNVAFNT-SHMHYNMINKWSRIADKIHGSKRLYE-LPRKAPHKVAS 1139

Qy        680 L---FCYFRRVTEKKPTGLVTFTRQSLEDFP-EWERCEKLLTRL----HVTYEGTIEGNG 731
Db       1140 LDVEACFLERTS-------------AFEQFPKQFSLVNKIVSKYKLQQNIYKEGYLLQDG 1186

Qy        732 QGMLQVDFANRFV---GGGVTSAGLVQEEIRFLINPELIV--------------SRLFTE 774
Db       1187 -GDLKGKIENRFFKLHGSQLSGYHEISRKAKIDINLLKVTKVLRNEDIQADNGGQRNFTD 1245

Qy        775 VLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALH---FR 831
Db       1246 WVLFNECFQLVFDD------GERITFNAECSNEEKS---DWYNKLQEVVELNVFHQPWVK 1296

Qy        832 RYLDQFVPEKIRR 844
Db       1297 KYCEKLAEEEKTR 1309
```

```
RESULT 5
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

  Query Match             3.4%;  Score 175;  DB 2;  Length 1359;
  Best Local Similarity   18.5%;  Pred. No. 0.014;
  Matches  141;  Conservative  124;  Mismatches  284;  Indels  214;  Gaps   32;

Qy         87 KGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQY 146
Db         81 KSRKRAKSES-ESDESDEEEDRKKSKSKKKVDQKKKEKSKKKRTTSSSEDEDSDEEREQ- 138

Qy        147 LKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESR 206
Db        139 -KSKKKSKKTKKQTSSESSE---ESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSK 194

Qy        207 DNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQEETDVVSESPLSD 266
Db        195 KSKKGLK--KKAKSESESESEDEKEVKKSKKKSKKVVKKESESEDEAPEKKKTEKRKRSK 252

Qy        267 TGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEP----ESPMDVDNSKNSCQDSEADEE 322
```

```
Db        253 TSSEE------------SSESEKSDEEEEEKESSPKPKKKKPLAVKKLSSDEESEESDVE 300

Qy        323 TSPG---------FDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEG 373

Db        301 VLPQKKKRGAVTLISDSEDEKDQKSESEASDVEEKVSKKKAKKQESSE----------SG 350

Qy        374 GESRAGMNDVNAKRPGSTSSLNVECRNSKQHG--RKDSKITDHFMRVPKAEDKRKEQCEM 431

Db        351 SDSSEGSITVNRK------SKKKEKPEKKKKGIIMDSSKLQKETIDAERAEKERRKRLEK 404

Qy        432 KHQR--------------------TERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIR 471

Db        405 KQKEFNGIVLEEGEDLTEMLTGTSSQRKLKSVV---LDPD----SSTVDEESKKP----- 452

Qy        472 LPPLRPSANHTVTIRVDLLRIGEVPKPFPTH-FKDLWDNKHVKMPCSEQNLYPVEDENGE 530

Db        453 -----------VEVHNSLVRI---LKPHQAHGIQFMYD-------CAFESLDRLDTEG-- 489

Qy        531 RAAGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHL 590

Db        490 -----------------------SGGILAHCMGLGKT---LQVITFLHTVL-------M 515

Qy        591 YQSILPDMVKIALCLP-NICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAK 649

Db        516 HEKIGEKCKRVLVVVPKNVIINWF----KEFQKWLVDNDEELDTIDVN----------- 559

Qy        650 MKSEYSSYPDINFNR--LFEGRSSRKPEKLKTLFCYFRRVT------EKKPTGLVTFTRQ 701

Db        560 ---ELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKKKKPKNRNRRLEK 616

Qy        702 SLEDF---------PEWERCE-----------------KLLTRLHVTYEGT-IEGN----- 730

Db        617 AKEDFRKYLQNPGPDMVVCDEAHKLKNDDSALSKCMVKILTKRRICLTGTPLQNNLMEYH 676

Qy        731 ------GQGML--QVDFANRFVGGGVTSAGLVQE----EIRFL 761

Db        677 CMVNFVKPGLLGTKTEFANRFV--NIINRGRTKDASPLEVSFM 717
```

RESULT 6
A47328
natural killer cell tumor-recognition protein - human
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumc
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C;Accession: A47328
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824; PMID:8421688
A;Accession: A47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1403 <AND>
A;Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252
A;Experimental source: NK killer cells from adult blood
A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)
C;Genetics:
A;Gene: GDB:NKTR
A;Cross-references: GDB:137171; OMIM:161565
A;Map position: 3p23-3p21
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>

```
  Query Match             3.2%;  Score 166;  DB 1;  Length 1403;
  Best Local Similarity   19.7%;  Pred. No. 0.053;
  Matches  136;  Conservative   97;  Mismatches  253;  Indels  204;  Gaps   28;

Qy         39 VLDSKDAPVQFRVPPSSSGC--ALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESES 96

Db        661 VIPLSDSPPPSRWKPGEKPWKPSYERIQEMKAKTTHLLPIQSTY-SLANIK--ETGSSSS 717
```

```
Qy         97 LHSKENNNTREESMMSSV---------------QKDNFYQHNMEKLENV-SQLGFDKSPV 140

Db        718 YHKREKNSESDQSTYSKYSDRSSESSPRSRSRSSRSRSYSRSYTRSRSLASSHSRSRSPS 777

Qy        141 EKG---TQYLKQHQTAAMCKW----QNEGPHSERLLESEPPAVTLVPEQFSNANV----- 188

Db        778 SRSHSRNKYSDHSQCSRSSSYTSISSDDGRRAKRRLRSSGKKNSVSHKKHSSSSEKTLHS 837

Qy        189 ------DQSSPKDDHSDTNS--EESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Db        838 KYVKGRDRSSCVRKYSESRSSLDYSSDSEQ--SSVQATQSAQEKE-KQGQMERTHNKQEK 894

Qy        241 ACHPAEACAGCQQEETDVVSESPLSD--TGSEDVGTGLKNANRLNRQESSLGNSPPFEKE 298

Db        895 ---------NRGEEKSKSERECPHSKKRTLKENLSDHLRNGSKPKRK-----NYAGSKWD 940

Qy        299 SEPESPMDV-DNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKR 357

Db        941 SESNSERDVTKNSKNDSHPSSDKEEGEATSDSESEVSEIHIKVKPT------------TK 988

Qy        358 SSAKGGEIRLHFQFEGGESRAGMNDVN-----AKRPGSTSSLNVECRNSK---------Q 403

Db        989 SS----------------TNTSLPDDNGAWKSSKQRTSTSDSEGSCSNSENNRGKPQKHK 1032

Qy        404 HGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMR 463

Db       1033 HGSKENLKREH---TKKVKEKLKGKKDKKHKAPKRKQAFHWQP----------------- 1072

Qy        464 RMPRCGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYP 523

Db       1073 --------------------------------PLEFGEEEEEEIDDKQVTQESKEKKV-- 1098

Qy        524 VEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVA----YSKKWDFTALIDFWD 579

Db       1099 --SENNE----------------------TIKDNILKTEKSSEEDLSGKHDTVTVSSDLD 1134

Qy        580 KVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAF 639

Db       1135 QFTKDDSKLSISPTALNTEENVA-CLQNI---------QHVEESVPNGVEDVLQTDDNME 1184

Qy        640 FCTFPRRNAKMKSEYSS--------YPDIN 661

Db       1185 ICT-PDRSSPAKVEETSPLGNARLDTPDIN 1213
```

RESULT 7
T19137
hypothetical protein C09G1.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19137
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19079
A;Accession: T19137
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1115 <WIL>
A;Cross-references: EMBL:Z50176; PIDN:CAA90539.1; GSPDB:GN00028; CESP:C09G1.2
A;Experimental source: clone C09G1
C;Genetics:
A;Gene: CESP:C09G1.2
A;Map position: X
A;Introns: 14/1; 49/2; 1095/2

```
  Query Match             3.2%;  Score 164;  DB 2;  Length 1115;
  Best Local Similarity   19.5%;  Pred. No. 0.051;
  Matches  167;  Conservative   98;  Mismatches  306;  Indels  286;  Gaps   39;

Qy          1 MSAGPGCEP--CTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGC 58

Db        106 ISKSTNMEPISCNSKIKLSQKQTSPQTDNDPDYFTAKPTDKITKR--------PSISRGT 157
```

```
Qy     59 ALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDN 118
Db    158 AM----QNEGS----LFHDKR----QNTLG---TEHETLKNTVKNT------HPKVDKD- 195
Qy    119 FYQHNMEKLENVSQLG-----FDKSP---VEKGTQYLKQ----HQTAAMCKWQNEGPHSE 166
Db    196 ---HKRIENQDISSKGQEPTCFQKKARLRTEIDTNYLPAKISIHSRE---NWTRDREHLR 249
Qy    167 RLLESEPP------------------------AVTLVPEQFSNANVDQSSPKDDHSDT 200
Db    250 RLLESEEPNPKKQKSVPTEHDLKKRGWHSSSTHQSMQRIQKQVSSEEVDSKNQKSAQSET 309
Qy    201 NSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQEETDVVS 260
Db    310 D----------LNYFTLNNNSSKTLFEQQRNVRTDQDLN---NNRDSCSACGNN------ 350
Qy    261 ESPLSDTGSEDVGTGLKNANRLNRQESS----------------LGNSPPFEKESEPESP 304
Db    351 ----SDAEKEDI-KGTRKRSSLENQHSERMQFSQTEKSSPIGDMVQNSPDF-KNMANRTP 404
Qy    305 MDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGE 364
Db    405 AESEKGKKKAQDWQHSLNFRSG----------------RSAINKQQNEIKIKPSSSNDQ 447
Qy    365 IRLHFQF---EGGESRAGMNDVNAKRPGSTSSLNVE--CRNSKQHGRKDSK---ITDHFM 416
Db    448 KR-HWLFKMRESAQTKKNLKYLIAKRTHQSMQCQSEFKCVSSSNQTEKDLKSFSIKKNVQ 506
Qy    417 RVP-----KAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIR 471
Db    507 VQPCSPSLKSSDIGNERNRIKRQ------------HFS-----LGEQSFEKRK------- 542
Qy    472 LPPLRPSANHTVTIRVDL--------------------LRIGEVPKPFPTHFKDLWDNK 510
Db    543 -PALTATGNEKGDLKRKLSLLKDQFSAKQTEQSLQCSADLEDGEI------------DEK 589
Qy    511 HVK-----------------MPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRPQN 553
Db    590 HCQILSNSQVNRVFKKRKTALAAKEPNNFTVSCCSDSNSTGITDDPIQT---------KN 640
Qy    554 LKDAILKYNVAYSKKWDFTALI-----------------DFWDKVLEEAEAQHLYQSILP 596
Db    641 TKLSLEEQDPLLSERRKTLATIFNQKKPVQIDQNLQCSFDLEDGEIDEKQCQISLHSQVH 700
Qy    597 DMVKIALCLP-----------------------NICTQPIPLLKQKMNHSITMSQEQIA 632
Db    701 RVLKKRQLAPAAKDLNNFTVNCSLDSNSNESNDEPIKTEKTQLLLKEHQPVISKRQETLK 760
Qy    633 SLLA-NAFFCTFPRRNAKMKSEYSSYPDINF-NRLFEGRSSRKPEKLKTLFC-------- 682
Db    761 TESSFNNFNSTDAHRNCRRQSKLIEKQPISLEDSVFEQNKSAQPESVINYFTIDNKNQDW 820
Qy    683 -YFRRVTEKKPTGLVTF 698
Db    821 QYFSG-PEKDPK-LVSF 835
```

RESULT 8
E85343
hypothetical protein AT4g29440 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: E85343
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: E85343
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1071 <STO>
A;Cross-references: GB:NC_001268; NID:g7269843; PIDN:CAB79702.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g29440
A;Map position: 4

Query Match 3.1%; Score 162.5; DB 2; Length 1071;
Best Local Similarity 21.5%; Pred. No. 0.06;
Matches 111; Conservative 76; Mismatches 194; Indels 135; Gaps 24;

```
Qy     42 SKDAPVQFRVPPSSSGCALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKE 101
Db    389 SRENSLEMR--PNDSFAKIGREKQQPG------MDDINLSSSADVLNKKQSSRASSHSPS 440
Qy    102 NNNTREESMMS-------SVQKDNFYQHNMEKLENVSQLGFDKSPV-------------- 140
Db    441 SNFSDDNDVTALDHIDSPSIFEENKFQSTVGDRES-----YNDSPVVVVAPAFDDYSSFF 495
Qy    141 ---------------EKGTQY--------LKQHQTAAMCKWQNEG---------PHSERL 168
Db    496 DKPQFDTEDAYHDEPEQGLGFSLLGSSSKTSDHMPTEISSWSLEGHKDLGKLSSASTSQV 555
Qy    169 LESEPPAVTLVPEQFSN------ANVDQSSP--KDDHSDTNSEESRDNQQFLTHVKLANA 220
Db    556 LEKEKPS---SPPTFDDGPTSPPASLHEPEPSAKFDDYDRDSESEEDNLGRLS--GRAEG 610
Qy    221 KQTMEDEQGREARSHQKCGKACHPA----EACAGCQQEETDVVSESPLSDTGSE--DVGT 274
Db    611 KSKLTAQKSHMSEGPDDLGRYFFPSDTEDQGDDSKTQEESD--AETP---TGLKFGPLAS 665
Qy    275 GLKNANRLNRQESSLGNSPPFEKESEPE----SPMDVDNSKNSCQDSEADEETSPGFDEQ 330
Db    666 GLENETTL----PSYGSSPPRDKTSSKSIKEYLPTEVDPSRSSSLQTASSSSIRNELYTQ 721
Qy    331 EDSSSAQTANKPSRFQP------READTELRKRSSAKGGEIRLHFQFEGGESRAGMN--D 382
Db    722 KASNSDK---RPSSIPPDSSSSDDESDMELPKRVSFRYQEKRTESRTRPTHLHSGVSHKD 778
Qy    383 VNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRT------ 436
Db    779 LEEEIPTRASTRSQDRRTHKTTPASASASYFHTMS-SDDEDEKEVHRDTAHIQTRPYISI 837
Qy    437 -------ER-------KIPKY-----IPPHLSPDKK 453
Db    838 SRRTKGQERRPSLVTAKIDKVSFDEESPPKLSPEAK 873
```

RESULT 9
S15901
chromogranin B precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S15901; S20727; A61076; S21773; S39369
R;Bauer, J.W.; Fischer-Colbrie, R.
Biochim. Biophys. Acta 1089, 124-126, 1991
A;Title: Primary structure of bovine chromogranin B deduced from cDNA sequence.
A;Reference number: S15901; MUID:91223091; PMID:2025642
A;Accession: S15901
A;Molecule type: mRNA
A;Residues: 1-646 <BIO>
A;Cross-references: EMBL:X55027; NID:g11; PIDN:CAA38846.1; PID:g12
R;Grandy, D.K.; Leduc, R.; Makam, H.; Flanagan, T.; Diliberto, E.J.; Thomas, G.; Civelli,
submitted to the EMBL Data Library, October 1990
A;Description: Nucleotide and deduced amino acid sequence of bovine adrenal medulla Chror
A;Reference number: S20727
A;Accession: S20727
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 21-63,'N',65-92,'FRS',96,'RA',99-100,'S',101-480,'L',482-646 <GRA>
A;Cross-references: EMBL:X55489; NID:g201; PIDN:CAA39109.1; PID:g202
R;Flanagan, T.; Taylor, L.; Poulter, L.; Viveros, O.H.; Diliberto Jr., E.J.
Cell. Mol. Neurobiol. 10, 507-523, 1990
A;Title: A novel 1745-Dalton pyroglutamyl peptide derived from chromogranin B is in the l
A;Reference number: A61076; MUID:91223542; PMID:1982622
A;Accession: A61076

A;Molecule type: protein
A;Residues: 567-580 <FLA>
R;Dillen, L.; Boel, S.; de Potter, W.P.; Claeys, M.
Biochim. Biophys. Acta 1120, 105-112, 1992
A;Title: Mass spectrometric characterization of bovine chromaffin granule peptides relat
A;Reference number: S21773; MUID:92207983; PMID:1554736
A;Accession: S21773
A;Molecule type: protein
A;Residues: 567-580 <DIL>
R;Yoo, S.H.
Biochim. Biophys. Acta 1179, 239-246, 1993
A;Title: pH-dependent binding of chromogranin B and secretory vesicle matrix proteins tc
A;Reference number: S39369; MUID:94032431; PMID:8218367
A;Accession: S39369
A;Molecule type: protein
A;Residues: 239-244;562-565 <YOO>
C;Superfamily: chromogranin B precursor
C;Keywords: adrenal gland; glycoprotein; pituitary; pyroglutamic acid; sulfoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-646/Product: chromogranin B #status predicted <MAT>
F;567-580/Product: pyroglutamyl peptide BAM-1745 #status experimental <MA2>
F;36-57/Disulfide bonds: #status predicted
F;188/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;315/Binding site: sulfate (Tyr) (covalent) #status predicted
F;567/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

```
  Query Match             3.1%;  Score 161.5;  DB 1;  Length 646;
  Best Local Similarity   20.7%;  Pred. No. 0.035;
  Matches  158;  Conservative   87;  Mismatches  249;  Indels  269;  Gaps   39;

Qy         15 RWDAAATSPPAASDARSFPGRQRRVLDS--KDAPVQFRVPPSSSGCALGRAGQHRGSATS 72
Db         80 RFEVRLLRDPA--DTSEAPGLSSRE-DSGEGDAQVPTVADTESGGHSRERAGEPPGSQVA 136

Qy         73 LVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQ 132
Db        137 ------------------KEAKTRYSKSEGQNREEEMVK-------YQ-KRERGEVGSE 169

Qy        133 LGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSS 192
Db        170 ERLSEGPGKAQTAFLNQ---------RNQTPAKKEELVSR-------YDTQSARGLEKSH 213

Qy        193 PKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQ 252
Db        214 SRERSSQESGEETKSQENW--------PQELQRHPEGQEA-----------PGES----- 249

Qy        253 QEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKN 312
Db        250 --EEDA---SPEVD-------------KRHSRPRHHHGRSRPDRSSQEGNPPL------- 284

Qy        313 SCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHPQFE 372
Db        285 -------EEESHVG----TGNSDEEKARHPAHFRALEEGAEY-------GEEVRRHSA-- 324

Qy        373 GGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEMK 432
Db        325 ------------AQAPGDLQGARFGGRGRGEH---------QALRRPSEES--LEQENKR 361

Qy        433 HQRTERKIPKYIPPHLSPDKKWLGTPIEEM--RRMPRCGIRLPPLR----PSANHTVTIR 486
Db        362 H------------GLSPDLNMAQGYSEESEEERGPAPG---PSYRARGGEAAAYSTLGQ 405

Qy        487 VDLLR-IGEVPKPFPTHFK---DLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELIQT 542
Db        406 TDEKRFLGE------THHRVQESQRDKARRRLPGELRNYLDYGEEKGEEAARGKWQ---- 455

Qy        543 ALLNRLTRPQ-NLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 601
Db        456 --------PQGDPRDA--------------------DENREEARLRG------------ 474

Qy        602 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFF-------CTFPRR-------- 646
Db        475 ---------------KQYAPHHIT--EKRLGELL-NPFYDPSQWKSSRFERKDPMDDSFL 516

Qy        647 -----NAKMKSEYSSYPDINFN----RLFE-----GRSSRKPEKLKTLFCYFRRVTEKKP 692
Db        517 EGEEENGLTLNEKNFFPEYNYDWWEKKPFEEDVNWGYEKRNPVPKLDLKRQYDRVAELDQ 576

Qy        693 TGLVTFTRQSLEDFPEWERCEKLLTRLHVTYEGTIEGNGQGML 735
Db        577 --LLHYRKKSAE-FPDFYDSEEQMSPQH-TAENEEEKAGQGVL 615
```

RESULT 10
T49042
hypothetical protein T5P19.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49042
R;Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25014
A;Accession: T49042
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1535 <BEN>
A;Cross-references: EMBL:AL163972; GSPDB:GN00061; ATSP:T5P19.60
A;Experimental source: cultivar Columbia; BAC clone T5P19
C;Genetics:
A;Gene: ATSP:T5P19.60
A;Map position: 3
A;Introns: 38/3; 96/1

```
  Query Match             3.1%;  Score 159;  DB 2;  Length 1535;
  Best Local Similarity   18.9%;  Pred. No. 0.16;
  Matches  182;  Conservative  133;  Mismatches  338;  Indels  312;  Gaps   47;

Qy          6 GCEPCTKRPRW-----DAAATSPPA--ASDARSFPGRQRRVLDSKDAPV-----QFRVPP 53
Db         89 GCDSILHAKRWEPDGNDHTNTIPEALLSSQNRSLSAEVESPEDGSRTPMRTTHREYNSRP 148

Qy         54 SSSGCALGRAGQHRGSATSLVFKQKTI---TSWMDTKG--IKTVESESLHSKENN--NTR 106
Db        149 STS--------VERGYHPETVYKPETSDIRREWMRRTDDFSETGDSDVFTSERSSPYNTR 200

Qy        107 EESMM---------------------------------------SSVQKDNFYQHNMEKL 127
Db        201 SNAAQWAQHEGRYADPPRVPFYPASPSPSSAYEYGYSSPFHGSYVSASEQSYYHQQPNQF 260

Qy        128 ENVSQLG-FDKSPVEKGTQY--------------------LKQHQTAAMCKWQNEGPH-- 164
Db        261 EQYSREGWFQESSVASPTRFPGETSDGKYYHRSSQSQLHDLQYHNLYEPSR--SETPHHS 318

Qy        165 --SERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEES--RDNQQFL-------- 212
Db        319 VYSER---SYVPAAAPHRSTYSEHSVGISK-----SDTSSEKSILRNKKRYVRERNPVVK 370

Qy        213 THV-------KLANAKQTME-------DEQGREARSHQKCGKACHPAEACAGCQ----QE 254
Db        371 RHILPSAGGAPFATCSYCLELLQLPQVSPQGKRQRYQVRCG-------SCSGVLKFSIRE 423

Qy        255 ETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKNSC 314
Db        424 KADTVLDSP----SFVDYGMDFADETVTNHQDSA----------SEGHEEINPDGSHLTC 469

Qy        315 QDSEADEETSPGFDEQEDSSSAQTANKP--SRFQPREADTELRKRSS----AKGGEIRLH 368
Db        470 LDDDSGETIC-------KSNDAVILSQKILETFEDKGIKEDIRNISTKFLDLKLEALQPH 522

Qy        369 FQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQ 428
Db        523 LKPMGNRRLSEQQPTSSETIGETSKIHLE-QSQEAHSEKSTEMDNNICDIEEPE------ 575

Qy        429 CEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVD 488
Db        576 ----YEKNEIIKPEEIVGEGSGES--LEVPVNQNERM----------------------- 606
```

```
Qy        489 LLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDEN-GERAAGSRWELIQTALLNR 547
                                             || :    |||  ||:     ||:
Db        607 ----------------------------SESS---EEDERVAERSVSHSEELLYKKEYVS 635

Qy        548 LTRPQ--NLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSI--LPD--MVKI 601
                || |   |   |          |  :|:|       :  || |  :    :  |    :  |
Db        636 ETRVQSENTGD-------TYEFEWEET-------QETEEYEGESTLEEIGDKPQLHLDKS 681

Qy        602 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 661
                    :||    :   :   :  |:     | |:: :::|          ::|    |  :|   :|
Db        682 RYTRDSICDSSLSYGEPFEDTSVKEDTEHISNTISDA-------KSATTHSPLNSL--VN 732

Qy        662 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCE-------- 713
              |: | |    :   |    |:       |: |           ::      |:   |  :   |
Db        733 ENQRFIGHLDQSGEATVTI----SRIQE--------HSKYEYENLSERQELEMTISDRVT 780

Qy        714 -KLLTRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLF 772
                ||     :  |  |  |:|  :   |                    | :  |    ||    :|   :    ||
Db        781 CKLEESRYETNE-TLEPSKVG----------EDGSLVSFEKVSEAFETIIEGKTGEERLV 829

Qy        773 TEVLDH-NECLIIT-GTEQ--YSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDAL 828
              :   :|   ||  :    :  || :    |  | :|:|     |   :  |||:                 |
Db        830 SHQMDSPNENVENSYGTLEPVYLEESGFAGERTWEGTIEDRA----------------GL 873

Qy        829 HFRRY 833
               |   |
Db        874 HLEEY 878
```

RESULT 11
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

```
  Query Match             3.1%;  Score 159;  DB 2;  Length 6642;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches  185;  Conservative  136;  Mismatches  387;  Indels  206;  Gaps   45;

Qy          9 PCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRAGQHRG 68
               |  |  |            ||| :  :|  | :: :    |   :|  :       |||      |      :  :
Db       1407 PKEKSPASPTKKEKSPAAEEVKS-PTKKEK---SPSSPTKKEKSPSSPTKKTGDEVKEKS 1462

Qy         69 SATSLVFKQKTITSWMDTK------------GIKTVESE-SLHSKENNNTRE-----ESM 110
                |     |:|:     |  |               | | ||  ::  |    |  |  |       |
Db       1463 PPKSPTKKEKSPEKPEDVKSPVKKEKSPDATNIVEVSSETTIEKTETTMTTEMTHESEES 1522

Qy        111 MSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLE 170
               :||:|:   :     || :: ::     ||||  :    |: :|                :  :  |
Db       1523 RTSVKKEKTPEKVDEKPKSPTKK--DKSPEKSITEEIKSPVK------KEKSPEKVEEKP 1574

Qy        171 SEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTME-DEQG 229
               :  |      ||:    |:   :  |    :    |    |   |:|  :   :    :  :|       |     :  |::
Db       1575 ASPTKKEKSPEK--PASPTKKSENEVKSPTKKEKSPE-KSVVEELKSPKEKSPEKADDKP 1631

Qy        230 REARSHQKCGKACHPAEACAGCQQEET-DVVSESPLSDTGSEDVGT-----GLKNANRLN 283
               :    :|   :      :    ::|:: :  |  |  |  |    |        |        :|:   :
Db       1632 KSPTKKEKSPEKSATEDVKSPTKKEKSPEKVEEKPTSPTKKESSPTKKTDDEVKSPTKKE 1691

Qy        284 RQESSLGNSP--PFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANK 341
               :    ::    |    |  :||  ||           |:   ::  :: :|  |  ||     ||::   |    :
Db       1692 KSPQTVEEKPASPTKKEKSPE--------KSVVEEVKSPKEKSPEKAEEKPKSPTKKEKS 1743

Qy        342 PSRFQPREADTELRK-RSSAKGGEIRLHFQFEGGESRAGMNDVNAKRP----GSTSSLNV 396
               | :     |   :   :|  :|    |   |  :         :       |   |    ||   ||       |      :
Db       1744 PEKSAAEEVKSPTKKEKSPEKSAEEKPKSPTKKESSPVKMADDEVKSPTKKEKSPEKVEE 1803

Qy        397 ECRNSKQHGRKDSKITDHFMRVP------------KAEDKRKEQC-EMKHQRTERKIPKY 443
               :   :   :   :     |         :: |                |  |:  ||:   |     :: :      ||
Db       1804 KPASPTKKEKTPEKSAAEELKSPTKKEKSPSSPTKKTGDESKEKSPEKPEEKPKSPTPKK 1863

Qy        444 IPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHF 503
                ||    ||   ||  ||        :|        |      :  |              | |           :  :|        ||             |
Db       1864 SPPG-SPKKKKSKSP--EAEKPP-----APKLTRDLKLQTVNKTDLAHF-EVVVEHATEC 1914

Qy        504 KDLWDNKHVKMPCSEQNLYPVEDENGE--------------------RAAGSRWELIQTA 543
               |    |  |  |:      :   |  :     :|:   |                         ||||      ::|
Db       1915 KWFLDGKEI---TTAQGVTVSKDDQFEFRCSIDTTMFGSGTVSVVASNAAGS----VETK 1967

Qy        544 L-LNRLTRPQNLKDAILKYNVAYSKKWDFT-ALIDF----WDKVLEEAEAQH--LYQ--- 592
                |    |    |:                         :||  :||  |   |        |   |   :   |   |   ||:
Db       1968 TELKVLETPKE------------TKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQ 2015

Qy        593 ------------SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFF 640
                                :|   :   |  :|  :||     |        :   : :: :    |   |   |  |
Db       2016 NGNLLEDGKNGVTIKNEENKSSLIIPN--AQDSGKITVEASNEVG-SSESSAQLTVNPPS 2072

Qy        641 CT-----FPR-------RNAKMKSEYSSYP------DINFNRLFEGRSSRKPEKLKTLFC 682
                |         |:          |:  |:   |  :|           ||    ::   |     ||      :||
Db       2073 TTPIVVDGPKSVTIKETETAEFKATISGFPAPTVKWTIN-EKIVE--ESRTITTIKTEDV 2129

Qy        683 YFRRVTEKK--PTGLVTFTRQSL--EDFPEWE-------RCEKLLTRL--HVTYEG---- 725
                | :::   |     ||  |  |   | |:     :|    :   :          :       |    ::|     |    ||
Db       2130 YTLKISNAKIEQTGTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLR 2189

Qy        726 ---TIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE-C 781
                     ::|     |                              |: :|:| : :         :|:||  :
Db       2190 WNLELDGPSPG-----------------------TEVSWLLNGQPLTKSDTVQVVDHGDGT 2227

Qy        782 LIITGTEQYSEYTG 795
                 :|    |     | :|
Db       2228 YHVTIAEAKPEMSG 2241
```

RESULT 12
T03277
pol protein - yeast (Candida albicans) retrotransposon pCal (fragment)
C;Species: Candida albicans
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 29-Oct-1999
C;Accession: T03277
R;Matthews, G.D.; Goodwin, T.J.D.; Butler, M.I.; Berryman, T.A.; Poulter, R.T.M.
J. Bacteriol. 179, 7118-7128, 1997
A;Title: PCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast Cand
A;Reference number: Z14877; MUID:98037512; PMID:9371461
A;Accession: T03277
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1576 <MAT>
A;Cross-references: EMBL:AF007776; NID:g2636718; PIDN:AAC49878.1; PID:g2636719
C;Genetics:
A;Mobile element: retrotransposon pCal

```
  Query Match             3.1%;  Score 158.5;  DB 2;  Length 1576;
  Best Local Similarity   19.2%;  Pred. No. 0.18;
  Matches  145;  Conservative  103;  Mismatches  267;  Indels  241;  Gaps   36;

Qy          5 PGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVL--DSKDAP-VQFRVPPSSSGCALG 61
               ||   | :|                   ||             |: :     :   :|||  :|    |     |
Db        660 PGDNPVSK---------PPQLGTETSVIGKSKEPITNHTKDAPSIQGRDHKRSESTA-- 707
```

```
Qy     62 RAGQHRGSATSLVFKQKTITSWMDTKGIKTVE---SESLHSKENNNTR---EESMMSSVQ 115
Db    708 QVGLSHQPQTGTPASEESKLSGTDHFGVDVVKETVSEDWHTSDYPETSAEDEQQNPSLSA 767

Qy    116 KDNFYQHNMEKLENVSQLGFD------KSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLL 169
Db    768 NKNRVTEKIDEGENISFPGGDDDSVVINSNVEQSN---VETEDAGNSPIQDEVSQEGRIL 824

Qy    170 ESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQG 229
Db    825 NEQTDIVDTVAKVIENEKISPINSLDDHTELATDSGNDSNSTESDIQSKNEISPVINEKN 884

Qy    230 REARSHQKCGKACHPAEACAGCQQEE---------------TDVVSESPLSDTGSE-DVG 273
Db    885 TEI--IQK-----HIESILADKRLDEFETYNVDEIENVINDDDIAEANPLPDENNDVQMN 937

Qy    274 TGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQEDS 333
Db    938 ESFDNNHSMSRAKKKY----TFEKEVNEK----IAGTKHSL------DTTDP-----REA 978

Qy    334 SSAQTANKPSRFQPREAD----TELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKR-- 387
Db    979 IRVLNTGETKRIEPKKREVPITVKLNKRSQYKSPYV----------TRSGRTVINPKRYL 1028

Qy    388 -----------PGSTSSLNVECRNSKQHGRKD-----------SKITDHFMRVPKAED-- 423
Db   1029 HAVVNKIDYNDPGWIKSMNAELEKFRS---KDVYEEVPIPTGVKPISMGWVHTEKIDSLK 1085

Qy    424 --KRKEQCEMKHQRTERKI----------------------------------------- 440
Db   1086 GVVRKSRCVVHGNRQKEKLDYDPFSVSSPVIDLVTIRLLTIIGCELGMTIQHLDVESAYL 1145

Qy    441 --------PKYI-PPHLSPDKK---WLGTPIEEMRR----MPRCGIRLPPLRPSANHTVT 484
Db   1146 NASITHSNPIYVFPPKSVPLKKNHCWL------LKRSVYGLKQSGLEW-------YHTIK 1192

Qy    485 IRV--------------------------------DLLRIGEVPKPFPTHFKDLWDNKH 511
Db   1193 -RVLEDIGFTQVLHNDGLFHIEYEEGSVIYLGLYVDDILMVGSSQKVIDNFVDQLRDHFE 1251

Qy    512 VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVAY--SKKW 569
Db   1252 VKVFGEISNYLGIEFRKTE----SGYILSQEKFLKKLLKDFKLDDSYGK-NIPWIPNDKY 1306

Qy    570 DFTALI--------DFWDKV----LEEAEAQHLYQS 593
Db   1307 EKVAIIRENVNPENDF-EKVPNETLLDPDAKKLYQS 1341
```

RESULT 13
A45988
dentin matrix acidic phosphoprotein AG1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Oct-1996
C;Accession: A45988
R;George, A.; Sabsay, B.; Simonian, P.A.; Veis, A.
J. Biol. Chem. 268, 12624-12630, 1993
A;Title: Characterization of a novel dentin matrix acidic phosphoprotein. Implications f
A;Reference number: A45988; MUID:93286101; PMID:8509401
A;Accession: A45988
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-489 <GEO>
A;Experimental source: odontoblast-pulp fibroblast, Sprague-Dawley
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:133593, NCBIP:133595)
C;Keywords: phosphoprotein

```
  Query Match             3.0%;  Score 158;  DB 2;  Length 489;
  Best Local Similarity   21.1%;  Pred. No. 0.039;
  Matches   78;  Conservative   72;  Mismatches  135;  Indels   84;  Gaps   17;

Qy     94 SESLHSKENNNTREESMMS--SVQKDNFYQHNMEKLENVSQLGFDKSPVEK-------GT 144
Db    122 ADTTQSSEDSTSQENSAQDTPSDSKD----HHSDEADSRPEAGDSTQDSESEEYRVGGGS 177

Qy    145 QYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPE---QFSNANVDQSSPKDDHSDTN 201
Db    178 EGESSHGDGS--EFDDEG------MQSDDPGSTRSDRGHTRMSSADISSEESKGDHEPTS 229

Qy    202 SEESRDNQ--QFLTHVKLANAKQTMEDEQG-------REARS------HQKCGKACHPAE 246
Db    230 TQDSDDSQDVEFSSRKSFRRSRVSEEDDRGELADSNSRETQSVSTEDFRSKEESRSETQE 289

Qy    247 ACAGCQQEETDVVSESPLSDTGSEDVG-----------TGLKNANRLNRQESSLGNSPPF 295
Db    290 DTAETQSQEDSPEGQDPSSES-SEEAGEPSQESSSESQEGVASESRGDNPDNTSQTGDQR 348

Qy    296 EKESEPESPMDVDNSKNSCQ-----DSEADEETSPGFDEQ-----EDSSS-----AQTAN 340
Db    349 DSESSEEDRLNTFSSSESQSTEEQGDSESNESLSLSEESQESAQDEDSSSQEGLQSQSAS 408

Qy    341 KPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSS------- 393
Db    409 RESRSQESQSEEDSRSEENRDS---------DSQDSSRSKEESNS--TGSTSSSEEDNHP 457

Qy    394 LNVECRNSK 402
Db    458 KNIEADNRK 466
```

RESULT 14
T18532
serine/threoine protein kinase - guinea pig
N;Alternate names: STE20-like kinase
C;Species: Cavia porcellus (guinea pig)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T18532
R;Itoh, S.; Kameda, Y.; Yamada, E.; Tsujikawa, K.; Mimura, T.; Kohama, Y.
Arch. Biochem. Biophys. 340, 201-207, 1997
A;Title: Molecular cloning and characterization of a novel putative STE20-like kinase ir
A;Reference number: Z18952; MUID:97288344; PMID:9143322
A;Accession: T18532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1231 <ITO>
A;Cross-references: EMBL:D88425; NID:g2911113; PIDN:BAA24930.1; PID:g2911114
A;Experimental source: strain Hartley; liver

```
  Query Match             3.0%;  Score 157.5;  DB 2;  Length 1231;
  Best Local Similarity   18.3%;  Pred. No. 0.15;
  Matches  147;  Conservative  117;  Mismatches  334;  Indels  205;  Gaps   29;

Qy     17 DAAATS---------PPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRAGQHR 67
Db    446 DHAVTSETNIEHNLKPEKERDQEKQPVLENKLVKSEDTTIQ------------------- 486

Qy     68 GSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKL 127
Db    487 ---TVDLVSQETGEKEVD---IHILDSEVVHAVEDTHEK-------LRKDDTTQKDV--I 531

Qy    128 ENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNAN 187
Db    532 SDTSSVG------ERDEEIGAVPKTAESSAEGAQGDGGKETDEGAQILISKATEGPKASG 585

Qy    188 VDQSSPKDDHSDTNSEESRDNQQFL---THVKL--ANAKQTMEDEQG---REARSHQKCG 239
Db    586 TEEAPPVTEITETN-----DTDQKLVENTHEKQLPISSETTLDTSEGLGASEGREVTESG 640

Qy    240 KACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKES 299
Db    641 ST-----------EEVEV--EGAVSETDEEDVQSETRGAPMAVTQMDTEKNETPHEAPA 686

Qy    300 E-----PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTEL 354
```

```
Db        687 QVEVQVPVPPQPSEPPPAPIPSININSEAAENKGEMGASLNTETILLPESESQKENDTDS 746

Qy        355 RKRSSAKGGEIRLHFQFEG-------------GESRAGMNDVNAKRPGSTSSLNVECRNS 401
                  |:|    | |:                 |:|     :    |        :|   |
Db        747 GTGSTADNSSIDLNLSISSFLSKTKDNGSISLQETRRQKKTLKKTRKFIVDGVEVSVTTS 806

Qy        402 KQHGRKDSKITD-HFMR---------VPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPD 451
                  |     |||  :   |:|         :  | |  :  ::|     | |:    :|  :      :
Db        807 KIVTDSDSKTEELRFLRRQELRELRFLQKEEQRAQQQLNGKLQQQREQIFRRFEQEMMSK 866

Qy        452 KKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRI-GEVPKPFPTHFKDLWDNK 510
              |:       || : :  :   |      |     || :|  :    || ||   |     :  |:::  |:
Db        867 KRQYDQEIENLEKQQKQTIE----RLEQEHTNRLRDEAKRIKGEQEKEL-SKFQNILKNR 921

Qy        511 HVKMPCSEQNLYPVEDENGERAAGSRWELIQT-----------------ALLNRLTRPQN 553
                  |    :|     | :|    |     : ||  |:                    | :: :  |
Db        922 K-KEVLNEVEKAP-KDLRKELMKRRKEELAQSQHVQEQDFVQKQQQELDGSLKKIIQQQK 979

Qy        554 LKDAILKYNVAYSKKWDFTAL-IDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQP 612
                : |  ::     :|:      |      |:   |||    |   :|
Db        980 AELANIERECLNNKQQLMRAREAAIWE--LEERHLQEKHQ-------------------- 1017

Qy        613 IPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSR 672
                 ||||::         | :  |:                :|:  |      |      ||  |    :|
Db       1018 --LLKQQLKDQYFMQRHQLL------------KRHEKETEQMQRYN----QRLIEELKNR 1059

Qy        673 KPE-----------KLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWE----------- 710
              : :             : ||    |::      |       |     :  :|   |    |
Db       1060 QTQERARLPKIQRSEAKTRMAMFKKSLRINSTATPDQDRDKIKQFSAQEEKRQKNERMAQ 1119

Qy        711 -------------RCEKLLTRLH 720
                           :||   :  ||
Db       1120 HQKHENQMRDLQLQCEANVRELH 1142
```

RESULT 15
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71623
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1192 <GAR>
A;Cross-references: GB:AE001373; GB:AE001362; NID:g3845097; PIDN:AAC71813.1; PID:g384509
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0115w

```
  Query Match             3.0%;  Score 157;  DB 2;  Length 1192;
  Best Local Similarity   19.8%;  Pred. No. 0.15;
  Matches   74;  Conservative   66;  Mismatches  171;  Indels   62;  Gaps   11;

Qy         84 MDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKG 143
               :|  |   :   |  |      || :    :|         ::     :    || :  ::     :|:   |||
Db        423 VDEKKTEKAEEELEEDKEESEKDKEESEKDKEESEKDKEESEKDKEKTEEDEEKTEDEKG 482

Qy        144 TQYLKQHQTAAMCKWQ---NEGPHSERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDT 200
              |:   |:         |  :          ||   |      |          |         |::   :   |    |
Db        483 TEVYKKETDVDEKKEKGEYGEGTDDEEDKEKE--------EDDEETKVEEKKTEKDEEGT 534

Qy        201 NSEESRDNQQFLTHVKLANAK------QTMEDEQGREARSHQKCGKACHPAEACAGCQQE 254
               :  ||    |:           |:      |      :|  |||:    |      :|  |          :|
Db        535 DYEEDTDDSDKDEETKVEEKKTERDEEETEEDEKETEVEK-KKTEK-----------DEE 582

Qy        255 ETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKNSC 314
               ||    :    :|    :||  | ::   :    :::|:   |        |::     |: :|  |    ::

Db        583 GTDYEED---TDDSDKDVETEVEETDAEDKEENEEGTDD--EEDKVEETDLD-DQEEDGE 636

Qy        315 QDSEADEETSPGFDEQEDSSSAQTANK-----PSRFQPREADTELRKRSSAKGGEIRLHF 369
              :|  |  |:|      |:::|      :   :|          :    :|  |   |   |   :  :
Db        637 EDKEDDKEKDKEDDKEDDKEKDKEDDKEKYKEDDKEDDKEDDKEKDKEDNKE-------- 688

Qy        370 QFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQC 429
                :   |      :     :   |          :     ::|:    ::|:|             :|  ||
Db        689 --KDKEDNKEKDKEDDKEKDKEDDKEKDKEDNKEKDKEDNK------------EKDKEDD 734

Qy        430 EMKHQRTERKIPK 442
              :  ||  :    |:|  |
Db        735 KEKHDKHVRRIKK 747
```

Search completed: May 26, 2004, 18:47:24
Job time : 33.8764 secs

OM protein - protein search, using sw model

```
Run on:          May 26, 2004, 17:20:49 ; Search time 15.6533 Seconds
                                          (without alignments)
                                          3249.964 Million cell updates/sec

Title:           US-09-302-812-2
Perfect score:   5184
Sequence:        1 MSAGPGCEPCTKRPRWDAAA..........YHAVESCTQTTNQPGQRTGA 977

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	180.5	3.5	1447	1	BUD4_YEAST	P47136 saccharomyc
2	174	3.4	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
3	166	3.2	1462	1	NKCR_HUMAN	P30414 homo sapien
4	162	3.1	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
5	161.5	3.1	646	1	SG1_BOVIN	P23389 bos taurus
6	159.5	3.1	513	1	DMP1_HUMAN	Q13316 homo sapien
7	159	3.1	6632	1	UN89_CAEEL	O01761 caenorhabdi
8	158.5	3.1	5596	1	MDN1_HUMAN	Q9nu22 homo sapien
9	156.5	3.0	773	1	ELA1_MOUSE	Q8cb77 mus musculu
10	156	3.0	3644	1	MINT_MOUSE	Q62504 mus musculu
11	156	3.0	3664	1	MINT_HUMAN	Q96t58 homo sapien
12	155	3.0	489	1	DMP1_RAT	P98193 rattus norv
13	154	3.0	406	1	SR40_YEAST	P32583 saccharomyc
14	153	3.0	913	1	IF38_HUMAN	Q99613 homo sapien
15	153	3.0	3122	1	DPOZ_MOUSE	Q61493 mus musculu
16	152.5	2.9	915	1	CE05_HUMAN	Q9nyf5 homo sapien
17	152	2.9	914	1	ORC1_YEAST	P54784 saccharomyc
18	152	2.9	1453	1	NKCR_MOUSE	P30415 mus musculu
19	149	2.9	503	1	DMP1_MOUSE	O55188 mus musculu
20	149	2.9	845	1	NFM_RAT	P12839 rattus norv
21	149	2.9	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
22	148	2.9	633	1	MLH_TETTH	P40631 tetrahymena
23	148	2.9	677	1	SG1_MOUSE	P16014 mus musculu
24	148	2.9	3259	1	GOB1_HUMAN	Q14789 homo sapien
25	147	2.8	2004	1	MYS3_HUMAN	Q92794 homo sapien
26	146.5	2.8	1210	1	AF4_HUMAN	P51825 homo sapien
27	146.5	2.8	1489	1	YGP0_YEAST	P53115 saccharomyc
28	145.5	2.8	500	1	GAR2_SCHPO	P41891 schizosacch
29	145.5	2.8	677	1	SG1_HUMAN	P05060 homo sapien
30	145.5	2.8	2224	1	FA5_HUMAN	P12259 homo sapien
31	145	2.8	950	1	STL2_MOUSE	Q99n50 mus musculu
32	144.5	2.8	668	1	SG1_PIG	Q9glg4 sus scrofa
33	144.5	2.8	757	1	CC5_SCHPO	P39964 schizosacch
34	144.5	2.8	1478	1	THO2_HUMAN	Q8ni27 homo sapien
35	144	2.8	598	1	CYL1_HUMAN	P35663 homo sapien
36	144	2.8	678	1	GARP_PLAFF	P13816 plasmodium
37	144	2.8	1362	1	BRD4_HUMAN	O60885 homo sapien
38	144	2.8	2459	1	MAPB_RAT	P15205 rattus norv
39	144	2.8	5560	1	SPEN_DROME	Q8sx83 drosophila
40	143.5	2.8	699	1	NB14_HUMAN	Q14978 homo sapien
41	143.5	2.8	1147	1	CARB_HUMAN	Q9bx17 homo sapien
42	143.5	2.8	1338	1	ACIN_MOUSE	Q9jix8 mus musculu
43	143.5	2.8	5120	1	PCLO_CHICK	Q9pu36 gallus gall
44	143	2.8	2492	1	ATRX_HUMAN	P46100 homo sapien
45	143	2.8	2492	1	ATRX_PONPY	Q7yqm3 pongo pygma

ALIGNMENTS

```
RESULT 1
BUD4_YEAST
ID   BUD4_YEAST     STANDARD;      PRT;  1447 AA.
AC   P47136;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Bud site selection protein BUD4.
GN   BUD4 OR YJR092W OR J1905.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96295555; PubMed=8707826;
RA   Sanders S.L., Herskowitz I.;
RT   "The BUD4 protein of yeast, required for axial budding, is localized
RT   to the mother/BUD neck in a cell cycle-dependent manner.";
RL   J. Cell Biol. 134:413-427(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ramezani Rad M., Kirchrath L., Hollenberg C.P.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Co-assembles with BUD3 at bud sites. BUD4 and BUD3 may
CC       cooperate to recognize a spatial landmark (the neck filaments)
CC       during mitosis and they subsequently become a landmark for
CC       establishing the axial budding pattern in G1.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- CAUTION: Ref.2 sequence differs from that shown in the N-terminal
CC       section due to frameshifts.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U41641; AAB17116.1; -.
DR   EMBL; Z49592; CAA89620.1; ALT_FRAME.
DR   EMBL; Z49591; CAA89619.1; -.
DR   PIR; S57113; S57113.
DR   GermOnline; 141924; -.
DR   SGD; S0003852; BUD4.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00169; PH; 1.
DR   SMART; SM00233; PH; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   ATP-binding.
FT   DOMAIN      181    189       ASP/GLU-RICH (ACIDIC).
FT   NP_BIND    1175   1182       ATP (POTENTIAL).
FT   DOMAIN     1302   1413       PH.
FT   CONFLICT    291    328       PTASVEDLNISTNLPGADSSQNNPVTTDADALIENDVV ->
FT                                ANSLCGGFKHFNESSRVLIPAKIIQSLLMRMRLLKTMLC
```

```
FT                                (IN REF. 2).
FT   CONFLICT    340    340       D -> E (IN REF. 2).
SQ   SEQUENCE   1447 AA;  164485 MW;  1D056B9FF1B7067B CRC64;

  Query Match             3.5%;  Score 180.5;  DB 1;  Length 1447;
  Best Local Similarity   16.8%;  Pred. No. 0.0099;
  Matches  184;  Conservative  156;  Mismatches  362;  Indels  391;  Gaps   45;

Qy           70 ATSLVFKQKTITSWMDTK----GIKTVESESLHSKEN-----NNTREESMMSSVQKDNFY 120
                :||   ::  |  :  |:      | :  || |: |           |   : :  :: :
Db          417 STSTKSEESYIADYKVTRQEDWDTKKLHQESEHANEQPAIIPQKDSSEETFTELNNESEF 476

Qy          121 QHNM------------------------EKLENVSQLGFD-------KSPVEKGTQYLKQ 149
                | |                         | : | |::| |         |: | :   ::
Db          477 QRNFKDGEEYRIVQHEESLYGQRTKSPEENIINGSEIGVDHGEAAEVNEPLAKTS--AEE 534

Qy          150 HQTAAMCKWQ--NEGPHSERLLE-------------------------SEPPAVTLVPE- 181
                |  :: |: |  :|  : :|: |                          ||  | |:|
Db          535 HDLSSSCEDQSVSEARNKDRIEEKEVETKDENIETEKDESEYHKVEENEEPEHVPLLPPL 594

Qy          182 ------QFSNANVDQSSPKDDHSD----------------------TNSEESRDNQQFLT 213
                      ||:   :|::   :|  |                      :|| ||  | ||
Db          595 PRWEEIQFNEPFIDENDTSNDSIDLTRSMKPSDYISIWHIQEEEIKSNSPESIANSQFSQ 654

Qy          214 HVKLANAKQTMEDEQGREARSHQ------KCGKACHPAEACAGCQ--QEETDVVSESP-- 263
                   :  |  |:: ::   : | :         :  :|    :       |  ::| |
Db          655 QSSITTA-STVDSKKDNGSTSFKFKPRIVSRSRIYNPKSRVSSLNYYDNEDYILSNSEWN 713

Qy          264 ----------LSDTGSEDVGTGLKNA-------NRLNRQESSLGNSPPFEKESEPESPMD 306
                          :|    ::: |   :|        :|| ::|   |        | | | :
Db          714 ALDPMRRNTLISKRIQDNIRTQKGHAPLIRPSIMKLNGEDSGFQN-----HFLEVEQPQE 768

Qy          307 VDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA------ 360
                 :|   |   || |  |: | |||:  :: |           |:   :|:  ||
Db          769 HENIPLSTHLSEQDITTNVGLDEQKLPTNTQD----------EAEISIREIESAGDITFN 818

Qy          361 KGGEIRLHFQFEGGESRAGMNDV----------------NAKRPGSTSSLNVECRNSKQ- 403
                :|  : | |  | |:  |   |                 : :|  |  | |    :| :
Db          819 RGDLLSLSFDEELGQDFANFLDALDHDSTSFNHGPDDSSSFQRDSSKKSFNSLWESSYEL 878

Qy          404 ----HGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTER-------------KIP----- 441
                      ||     |   ::   |   |:  :::  | ||             | |
Db          879 KPPPSIRKQPIAPDVLQKL--LESDTKDDADLEKIREERITEPRTGLGIGMLKTPVKDVS 936

Qy          442 -----------------------------------------KYIPP----HLSPDKKWLG 456
                                                         |  |      :|| |: :
Db          937 IALAASIKGYEASFSDTDSRPEGMNNSDAITLNMFDDFEEDKMTPSTPVRSISPIKRHVS 996

Qy          457 TP----------------------IEEMRRMPRCGIR--LPPLRPSANHTVTIRVDLLRI 492
                :|                      || | :    |::  :|||       |  : : :
Db          997 SPFKVVKAGNKQENNEINIKAEEEIEPMTQQETDGLKQDIPPLLAQTKDNVEAKEETITQ 1056

Qy          493 GEVPKPFPTHFKD---------------LWDNKHVKMPCSEQNLYPVEDENGERAAGSRW 537
                 | |:     | |               |:  |      | :  | :  :|||    : |
Db         1057 LEEPQDVEQEFPDMGTLYLSIKAISTLALYGTK------SHRATYAIVFDNGENVVQTPW 1110

Qy          538 ELI--------------------QTALLNRLTRPQNLKDAILKYNVAYSK-KWDFTALID 576
                | :                    :    :  :   : |  ::     | | : :   ::|
Db         1111 ESLPYDGNIRINKEFELPIDFKGKAETSSASSERDSYKKCVITLKCKYEKPRHELVEIVD 1170

Qy          577 ---------------FWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMN 621
                               |  | :::   |  :  :       | |
Db         1171 KVPVGKSFFGKTKYKFEKKYVQKKPKQDEWDYLFAQDGSFARC----------------- 1213

Qy          622 HSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPD--INFNRLFEGRSSRKPEKLKT 679
                  | :::|     | |  | |    :  | :::|   |      ||:|    : | |: :
Db         1214 -EIEINEE----FLKNVAFNT-SHMHYNMINKWSRIADKIHGSKRLYE-LPRKAPHKVAS 1266

Qy          680 L---FCYFRRVTEKKPTGLVTFTRQSLEDFP-EWERCEKLLTRL----HVTYEGTIEGNG 731
                |    |:  | :             : | || ::    |::::     ::  || :  :|
Db         1267 LDVEACFLERTS-------------AFEQFPKQFSLVNKIVSKYKLQQNIYKEGYLLQDG 1313

Qy          732 QGMLQVDFANRFV---GGGVTSAGLVQEEIRFLINPELIV--------------SRLFTE 774
                 | |:    |||    |  ::    :  : :  ||   :                |  ||:
Db         1314 -GDLKGKIENRFFKLHGSQLSGYHEISRKAKIDINLLKVTKVLRNEDIQADNGGQRNFTD 1372

Qy          775 VLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALH---FR 831
                 :  |||  :   :      |   |:    |:|::|   ||  :  |:| ::  |    :
Db         1373 WVLFNECFQLVFDD------GERITFNAECSNEEKS---DWYNKLQEVVELNVFHQPWVK 1423

Qy          832 RYLDQFVPEKIRR 844
                :| ::   |:  |
Db         1424 KYCEKLAEEEKTR 1436


RESULT 2
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   PIR; T34036; T34036.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;
KW   DNA-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
```

```
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;

  Query Match             3.4%;  Score 174;  DB 1;  Length 1359;
  Best Local Similarity   18.5%;  Pred. No. 0.022;
  Matches  141;  Conservative  124;  Mismatches  284;  Indels  214;  Gaps   32;

Qy         87 KGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQY 146
Db         81 KSRKRAKSES-ESDESDEEEDRKKSKSKKKVDQKKKEKSKKKRTTSSSEDEDSDEEREQ- 138

Qy        147 LKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESR 206
Db        139 -KSKKKSKKTKKQTSSESSE---ESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSK 194

Qy        207 DNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQEETDVVSESPLSD 266
Db        195 KSKKGLK--KKAKSESESESEDEKEVKKSKKKSKKVVKKESESEDEAPEKKKTEKRKRSK 252

Qy        267 TGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEP----ESPMDVDNSKNSCQDSEADEE 322
Db        253 TSSEE------------SSESEKSDEEEEEKESSPKPKKKKPLAVKKLSSDEESEESDVE 300

Qy        323 TSPG---------FDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEG 373
Db        301 VLPQKKKRGAVTLISDSEDEKDQKSESEASDVEEKVSKKKAKKQESSE----------SG 350

Qy        374 GESRAGMNDVNAKRPGSTSSLNVECRNSKQHG--RKDSKITDHFMRVPKAEDKRKEQCEM 431
Db        351 SDSSEGSITVNRK------SKKKEKPEKKKKGIIMDSSKLQKETIDAERAEKERRKRLEK 404

Qy        432 KHQR--------------------TERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIR 471
Db        405 KQKEFNGIVLEEGEDLTEMLTGTSSQRKLKSVV---LDPD----SSTVDEESKKP----- 452

Qy        472 LPPLRPSANHTVTIRVDLLRIGEVPKPFPTH-FKDLWDNKHVKMPCSEQNLYPVEDENGE 530
Db        453 -----------VEVHNSLVRI---LKPHQAHGIQFMYD-------CACESLDRLDTEG-- 489

Qy        531 RAAGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHL 590
Db        490 -----------------------SGGILAHCMGLGKT---LQVITFLHTVL-------M 515

Qy        591 YQSILPDMVKIALCLP-NICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAK 649
Db        516 HEKIGEKCKRVLVVVPKNVIINWF----KEFQKWLVDNDEELDTIDVN------------ 559

Qy        650 MKSEYSSYPDINFNR--LFEGRSSRKPEKLKTLFCYFRRVT------EKKPTGLVTFTRQ 701
Db        560 ---ELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKKKKPKNRNRRLEK 616

Qy        702 SLEDF--------PEWERCE-----------------KLLTRLHVTYEGT-IEGN----- 730
Db        617 AKEDFRKYLQNPGPDMVVCDEAHKLKNDDSALSKCMVKILTKRRICLTGTPLQNNLMEYH 676

Qy        731 ------GQGML--QVDFANRFVGGGVTSAGLVQE----EIRFL 761
Db        677 CMVNFVKPGLLGTKTEFANRFV--NIINRGRTKDASPLEVSFM 717

RESULT 3
NKCR_HUMAN
ID   NKCR_HUMAN     STANDARD;      PRT;  1462 AA.
AC   P30414;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE   related protein) (NK-TR protein).
GN   NKTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93133824; PubMed=8421688;
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS.
RA   Anderson S.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Component of a putative tumor-recognition complex.
CC       Involved in the function of NK cells.
CC   -!- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC       via its N-terminus.
CC   -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04288; AAA35734.2; -.
DR   EMBL; AF184110; AAD56402.1; -.
DR   PIR; A47328; A47328.
DR   HSSP; Q27450; 1A33.
DR   Genew; HGNC:7833; NKTR.
DR   MIM; 161565; -.
DR   GO; GO:0016018; F:cyclosporin A binding; TAS.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
FT   DOMAIN        1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN      219    240       ARG/LYS-RICH (BASIC).
FT   DOMAIN      421    457       ARG/LYS-RICH (BASIC).
FT   DOMAIN      970   1010       ARG/LYS-RICH (BASIC).
FT   DOMAIN      194    244       ARG/SER-RICH.
FT   DOMAIN      466    574       ARG/SER-RICH.
FT   DOMAIN      664    814       ARG/SER-RICH.
FT   DOMAIN     1311   1348       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE   1462 AA;  165676 MW;  D98A1147763EF527 CRC64;

  Query Match             3.2%;  Score 166;  DB 1;  Length 1462;
  Best Local Similarity   19.7%;  Pred. No. 0.068;
  Matches  136;  Conservative   97;  Mismatches  253;  Indels  204;  Gaps   28;

Qy         39 VLDSKDAPVQFRVPPSSSGC--ALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESES 96
Db        607 VIPLSDSPPPSRWKPGQKPWKPSYERIQEMKAKTTHLLPIQSTY-SLANIK--ETGSSSS 663

Qy         97 LHSKENNNTREESMMSSV---------------QKDNFYQHNMEKLENV-SQLGFDKSPV 140
Db        664 YHKREKNSESDQSTYSKYSDRSSESSPRSRSRSSRSRSYSRSYTRSRSLASSHSRSRSPS 723

Qy        141 EKG---TQYLKQHQTAAMCKW----QNEGPHSERLLESEPPAVTLVPEQFSNANV----- 188
Db        724 SRSHSRNKYSDHSQCSRSSSYTSISSDDGRRAKRRLRSSGKKNSVSHKKHSSSSEKTLHS 783

Qy        189 ------DQSSPKDDHSDTNS--EESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240
Db        784 KYVKGRDRSSCVRKYSESRSSLDYSSDSEQ--SSVQATQSAQEKE-KQGQMERTHNKQEK 840

Qy        241 ACHPAEACAGCQQEETDVVSESPLSD--TGSEDVGTGLKNANRLNRQESSLGNSPPFEKE 298
Db        841 ---------NRGEEKSKSERECPHSKKRTLKENLSDHLRNGSKPKRK-----NYAGSKWD 886
```

```
Qy        299 SEPESPMDV-DNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKR 357
              || |  || ||||    |  ||     | : : |      ||:              :
Db        887 SESNSERDVTKNSKNDSHPSSDKEEGEATSDSESEVSEIHIKVKPT-----------TK 934

Qy        358 SSAKGGEIRLHFQFEGGESRAGMNDVN-----AKRPGSTSSLNVECRNSK---------Q 403
              ||              :   : | |     :|:  |||     | ||:          :
Db        935 SS---------------TNTSLPDDNGAWKSSKQRTSTSDSEGSCSNSENNRGKPQKHK 978

Qy        404 HGRKDSKITDHFMRVPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMR 463
              || |::   :|     | ::| |  : : ||:  :||   :   |
Db        979 HGSKENLKREH---TKKVKEKLKGKKDKKHKAPKRKQAFHWQP----------------- 1018

Qy        464 RMPRCGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYP 523
                                              |  |    ::  |:|  |      |: :
Db       1019 -------------------------------PLEFGEEEEEEIDDKQVTQESKEKKV-- 1044

Qy        524 VEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVA----YSKKWDFTALIDFWD 579
                  || |                       :|| |||   :     | | |   :   |
Db       1045 --SENNE---------------------TIKDNILKTEKSSEEDLSGKHDTVTVSSDLD 1080

Qy        580 KVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAF 639
              :  ::     :  : |     :| || ||         | : .|:     | :     |
Db       1081 QFTKDDSKLSISPTALNTEENVA-CLQNI---------QHVEESVPNGVEDVLQTDDNME 1130

Qy        640 FCTFPRRNAKMKSEYSS--------YPDIN 661
                || | |::  | | :|         ||||
Db       1131 ICT-PDRSSPAKVEETSPLGNARLDTPDIN 1159
```

```
RESULT 4
PCLO_HUMAN
ID   PCLO_HUMAN     STANDARD;      PRT;  5147 AA.
AC   Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragments).
GN   PCLO OR ACZ OR KIAA0559.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 4405-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9Y6V0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC                                  VSP_003926, VSP_003927;
CC         Note=No experimental confirmation available;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19188; CAB60727.1; -.
DR   EMBL; AC004903; AAD20936.1; -.
DR   EMBL; AC004886; AAD21789.1; -.
DR   EMBL; AB011131; BAA25485.1; -.
DR   EMBL; BC001304; AAH01304.1; -.
DR   EMBL; AC004082; AAB97937.1; -.
DR   PIR; T00634; T00634.
DR   HSSP; P04410; 1A25.
DR   Genew; HGNC:13406; PCLO.
DR   MIM; 604918; -.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0045202; C:synaptic junction; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
```

```
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
KW   Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   NON_TER        1      1
FT   DOMAIN       400    465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                 P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING      499    523       C4-TYPE (POTENTIAL).
FT   ZN_FING      969    992       C4-TYPE (POTENTIAL).
FT   NON_CONS    1010   1011
FT   DOMAIN      2300   2325       POLY-PRO.
FT   DOMAIN      4391   4442       PDZ.
FT   DOMAIN      4544   4633       C2 DOMAIN 1.
FT   DOMAIN      5031   5121       C2 DOMAIN 2.
FT   VARSPLIC    4404   4404       S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                                 QTGKLMEG (in isoform 2).
FT                                 /FTId=VSP_003923.
FT   VARSPLIC    4534   4534       K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
FT                                 /FTId=VSP_003924.
FT   VARSPLIC    4576   4576       G -> GQVMVVQNAS (in isoform 2).
FT                                 /FTId=VSP_003925.
FT   VARSPLIC    4757   4761       TAHKS -> SKRRK (in isoform 2).
FT                                 /FTId=VSP_003926.
FT   VARSPLIC    4762   5147       Missing (in isoform 2).
FT                                 /FTId=VSP_003927.
SQ   SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

  Query Match             3.1%;  Score 162;  DB 1;  Length 5147;
  Best Local Similarity   18.4%;  Pred. No. 0.59;
  Matches  115;  Conservative  109;  Mismatches  225;  Indels  176;  Gaps   26;

Qy        105 TREESMMSSVQKDNFYQH-------NMEKLEN---------VSQLGFDKSPVEKGTQYLK 148
              | |: ::  ::||:| |          :: |||:           | |  :||  || ||
Db       1278 TDEKDILKGLKKDSFSQESSPSSPSDLAKLESTVLSILEAQASTLADEKS--EKKTQ--- 1332

Qy        149 QHQTAAMCKWQNEGPHSERLLE--SEPPAVTLVPEQFSNANVDQSS--PKDDHSD-TNSE 203
               |: :         | |  :    :  ||   :|:  |:   :   ::      ||    |   :|:
Db       1333 PHEVSP------EQPKDQEKTQSLSETLEITISEEEIKESQEERKDTFKKDSQQDIPSSK 1386

Qy        204 ESRDNQQFLTHVKLA-----NAKQTMEDEQG----REARSHQKCGKACHPAEACAGCQQE 254
              : ::  :|:  :         : :::  ||       |: |:       :     :   :    |
Db       1387 DHKEKSEFVDDITTRREPYDSVEESSESENSPVPQRKRRTSVGSSSSDEYKQEDSQGSGE 1446

Qy        255 ETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPE------------ 302
              | | : :  :   :   ||  :| ::    :   |     : :|    :|: | :
Db       1447 EEDFIRKQIIEMSADED-ASGSEDDEFIRNQLKEISSSTESQKKEETKGKGKITAGKHRR 1505

Qy        303 ------SPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPS------------R 344
                     : :| |  :        | ||      |||  :     :| :: |                  |
Db       1506 LTRKSSTSIDEDAGRRHSWHDEDDE----AFDESPELKYRETKSQESEELVVTGGGGLRR 1561

Qy        345 FQPREADTELRKRSSAKGGEIRLHFQF---------------------EGGESRAGMNDV 383
              |:  |  :: :   :  ||:   :  :       |                    ||  |      :
Db       1562 FKTIELNSTIADKYSAESSQKKTSLYFDEEPELEMESLTDSPEDRSRGEGSSSLHASSFT 1621

Qy        384 NAKRPGSTSSLNVECRNSKQHGRKDSK------ITDHFMRVPKAEDK------RKEQCEM 431
                | | |||: :   :|   | : :||          |     :|   ||          |:|:   :
Db       1622 PGTSPTSVSSLDEDSDSSPSHKKGESKQQRKARHRPHGPLLPTIEDSSEEEELREEEELL 1681

Qy        432 KHQRTERKI---------------------------PKYIPPHLSPDKKWLGTPIEEMRR 464
              | |   :|:|                              || |  :||| :     :| ||:|:
Db       1682 KEQEKQREIEQQQRKSSSKKSKKDKDELRAQRRRERPKTPPSNLSPIED--ASPTEELRQ 1739

Qy        465 MPRCGIRLPPLRPS--ANHTVTIRVDLLRIGEVPKPFPTHFKDLWD-NKHVKMPCSEQNL 521
                     : |  |  : :: :| |          |: |    : : :   |   |:| :
Db       1740 ----AAEMEELHRSSCSEYSPSIESD-------PEGFEISPEKIIEVQKVYKLPTAVSLY 1788

Qy        522 YPVEDENGERAAGSRWELIQTALLNRLTRPQNLKDAILKYNVAYSKKWDFTAL--IDFWD 579
               | :::: :  ||:                          : || |    |     |    :  |    :  |
Db       1789 SPTDEQSIMQKEGSQ--------------KALKSAEEMYEEMMHKTHKYKAFPAANERD 1833

Qy        580 KVLEEAE-------AQHLYQSILPD 597
                      :| |:              ::|:|:: |
Db       1834 EVFEKEPLYGGMLIEDYIYESLVED 1858

RESULT 5
SG1_BOVIN
ID   SG1_BOVIN      STANDARD;      PRT;   646 AA.
AC   P23389; O02707;
DT   01-NOV-1991 (Rel. 20, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Secretogranin I precursor (SgI) (Chromogranin B) (CgB) [Contains: GAWK
DE   peptide; Secretolytin].
GN   CHGB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal chromaffin;
RX   MEDLINE=91223091; PubMed=2025642;
RA   Bauer J.W., Fischer-Colbrie R.;
RT   "Primary structure of bovine chromogranin B deduced from cDNA
RT   sequence.";
RL   Biochim. Biophys. Acta 1089:124-126(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal medulla;
RX   MEDLINE=97282588; PubMed=9136897;
RA   Yoo S.H., Kang Y.K.;
RT   "Identification of the secretory vesicle membrane binding region of
RT   chromogranin B.";
RL   FEBS Lett. 406:259-262(1997).
RN   [3]
RP   SEQUENCE OF 21-646 FROM N.A.
RC   TISSUE=Adrenal medulla;
RA   Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J.,
RA   Thomas G., Civelli O., Viveros O.H.;
RL   Submitted (OCT-1990) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 634-646.
RC   TISSUE=Adrenal chromaffin;
RX   MEDLINE=95262699; PubMed=7744058;
RA   Strub J.-M., Garcia-Sablone P., Lonning K., Taupenot L., Hubert P.,
RA   van Dorsselaer A., Aunis D., Metz-Boutigue M.-H.;
RT   "Processing of chromogranin B in bovine adrenal medulla.
RT   Identification of secretolytin, the endogenous C-terminal fragment of
RT   residues 614-626 with antibacterial activity.";
RL   Eur. J. Biochem. 229:356-368(1995).
RN   [5]
RP   CHARACTERIZATION OF SECRETOLYTIN.
RX   MEDLINE=96184581; PubMed=8603705;
RA   Strub J.M., Hubert P., Nullans G., Aunis D., Metz-Boutigue M.-H.;
RT   "Antibacterial activity of secretolytin, a chromogranin B-derived
RT   peptide (614-626), is correlated with peptide structure.";
RL   FEBS Lett. 379:273-278(1996).
CC   -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule
CC       protein, which may be the precursor for other biologically active
CC       peptides. The 16 pairs of basic AA distributed throughout its
CC       sequence may be used as proteolytic cleavage sites.
CC   -!- FUNCTION: Secretolytin has antibacterial activity.
CC   -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory
CC       granules.
CC   -!- PTM: O-glycosylated (Probable).
CC   -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein
CC       family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X55027; CAA38846.1; -.
DR   EMBL; U88551; AAC48720.1; -.
DR   EMBL; X55489; CAA39109.1; -.
DR   PIR; S15901; S15901.
DR   InterPro; IPR001819; Chromogranin_AB.
DR   InterPro; IPR001990; Granin.
DR   Pfam; PF01271; Granin; 1.
DR   PRINTS; PR00659; CHROMOGRANIN.
DR   PROSITE; PS00422; GRANINS_1; 1.
DR   PROSITE; PS00423; GRANINS_2; 1.
KW   Sulfation; Cleavage on pair of basic residues; Signal.
FT   SIGNAL         1     20
FT   CHAIN         21    646       SECRETOGRANIN I.
FT   PEPTIDE      418    484       GAWK PEPTIDE.
FT   PEPTIDE      634    646       SECRETOLYTIN.
FT   DISULFID      36     57       BY SIMILARITY.
FT   MOD_RES      158    158       SULFATION (POTENTIAL).
FT   MOD_RES      315    315       SULFATION (BY SIMILARITY).
FT   CONFLICT      64     64       N -> S (IN REF. 1).
FT   CONFLICT      70     70       N -> D (IN REF. 2).
FT   CONFLICT      93     98       SEAPGL -> FRSPRAS (IN REF. 3).
FT   CONFLICT     181    181       T -> M (IN REF. 2).
FT   CONFLICT     261    261       H -> R (IN REF. 2).
FT   CONFLICT     386    386       P -> R (IN REF. 2).
FT   CONFLICT     481    481       H -> L (IN REF. 3).
FT   CONFLICT     597    597       M -> V (IN REF. 2).
SQ   SEQUENCE   646 AA;  73339 MW;  420DB1178FD9E415 CRC64;

  Query Match             3.1%;  Score 161.5;  DB 1;  Length 646;
  Best Local Similarity   20.7%;  Pred. No. 0.043;
  Matches  158;  Conservative   87;  Mismatches  249;  Indels  269;  Gaps   39;

Qy         15 RWDAAATSPPAASDARSFPGRQRRVLDS--KDAPVQFRVPPSSSGCALGRAGQHRGSATS 72
              |::        || |    ||   |  ||   || |            | | :  |||:  ||   :
Db         80 RFEVRLLRDPA--DTSEAPGLSSRE-DSGEGDAQVPTVADTESGGHSRERAGEPPGSQVA 136

Qy         73 LVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQ 132
                                |::: :||      ||| |:          ||     |:  |   |:
Db        137 --------------------KEAKTRYSKSEGQNREEEMVK-------YQ-KRERGEVGSE 169

Qy        133 LGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSS 192
                 : |  :  | :| |               :|:  |   :|            :  |     :::|
Db        170 ERLSEGPGKAQTAFLNQ---------RNQTPAKKEELVSR-------YDTQSARGLEKSH 213

Qy        193 PKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQ 252
                 ::   |  : ||::  :  :              ::       :|:||             ||:
Db        214 SRERSSQESGEETKSQENW--------PQELQRHPEGQEA-----------PGES----- 249

Qy        253 QEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKN 312
                 |  |     ||  |                      | :|      |  |  |      |   |:
Db        250 --EEDA---SPEVD------------KRHSRPRHHHGRSRPDRSSQEGNPPL------- 284

Qy        313 SCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFE 372
                       :||:    |      :|   :  |  |:  |:    |     |           | |:|  |
Db        285 -------EEESHVG----TGNSDEEKARHPAHFRALEEGAEY-------GEEVRRHSA-- 324

Qy        373 GGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEMK 432
                                |: ||             :|             :|  |    |    ||    :
Db        325 -----------AQAPGDLQGARFGGRGRGEH---------QALRRPSEES--LEQENKR 361

Qy        433 HQRTERKIPKYIPPHLSPDKKWLGTPIEEM--RRMPRCGIRLPPLR----PSANHTVTIR 486
              |                ||||          ||     |   |   |    |   |        :| ::   :
Db        362 H------------GLSPDLNMAQGYSEESEEERGPAPG---PSYRARGGEAAAYSTLGQ 405

Qy        487 VDLLR-IGEVPKPFPTHFK---DLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELIQT 542
                 |   | :||          || :        |        ::|     :|       :| || ||   :|:
Db        406 TDEKRFLGE------THHRVQESQRDKARRRLPGELRNYLDYGEEKGEEAARGKWQ---- 455

Qy        543 ALLNRLTRPQ-NLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 601
```

```
                           || : :||                      |:  |||   :
Db        456 --------PQGDPRDA-------------------DENREEARLRG----------- 474

Qy        602 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFF-------CTFPRR-------- 646
                               ||    | ||   ::::  || |  |:          | |:
Db        475 ---------------KQYAPHHIT--EKRLGELL-NPFYDPSQWKSSRFERKDPMDDSFL 516

Qy        647 -----NAKMKSEYSSYPDINFN----RLFE-----GRSSRKPEKLKTLFCYFRRVTEKKP 692
                     |     :|  : :|: |::     : ||      |     | |      |    : || |
Db        517 EGEEENGLTLNEKNFFPEYNYDWWEKKPFEEDVNWGYEKRNPVPKLDLKRQYDRVAELDQ 576

Qy        693 TGLVTFTRQSLEDFPEWERCEKLLTRLHVTYEGTIEGNGQGML 735
                |: : ::| | ||::   |: ::  | | |    |   |||:|
Db        577 --LLHYRKKSAE-FPDFYDSEEQMSPQH-TAENEEEKAGQGVL 615


RESULT 6
DMP1_HUMAN
ID   DMP1_HUMAN     STANDARD;      PRT;   513 AA.
AC   Q13316; O43265;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE   protein-1) (DMP-1).
GN   DMP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Molar;
RX   MEDLINE=97321043; PubMed=9177774;
RA   Hirst K.L., Simmons D., Feng J., Aplin H., Dixon M.J., McDougall M.;
RT   "Elucidation of the sequence and the genomic organization of the human
RT   dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the
RT   locus from a causative role in the pathogenesis of dentinogenesis
RT   imperfecta type II.";
RL   Genomics 42:38-45(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   McDougall M., Juan X., Simmons D., Feng J.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.
RX   MEDLINE=96163890; PubMed=8586437;
RA   Aplin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;
RT   "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1)
RT   to the dentinogenesis imperfecta type II critical region at chromosome
RT   4q21.";
RL   Genomics 30:347-349(1995).
CC   -!- FUNCTION: Could be involved in the induction of mineralization of
CC       extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q13316-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q13316-2; Sequence=VSP_004191;
CC   -!- TISSUE SPECIFICITY: Expressed in tooth particularly in
CC       odontoblast, ameloblast and cementoblast.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U89012; AAC51332.1; -.
DR   EMBL; U34037; AAA97602.1; -.
```

```
DR   EMBL; U65378; AAB87728.1; -.
DR   Genew; HGNC:2932; DMP1.
DR   MIM; 600980; -.
DR   GO; GO:0005578; C:extracellular matrix; TAS.
DR   GO; GO:0005509; F:calcium ion binding; TAS.
DR   GO; GO:0005178; F:integrin binding; TAS.
DR   GO; GO:0001503; P:ossification; TAS.
KW   Biomineralization; Extracellular matrix; Signal; Alternative splicing.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    513       DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    467    467       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        364    366       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC     46     61       Missing (in isoform 2).
FT                                /FTId=VSP_004191.
FT   CONFLICT     69     69       S -> C (IN REF. 2).
SQ   SEQUENCE   513 AA;  55782 MW;  2C1FDE319A5D106F CRC64;

  Query Match             3.1%;  Score 159.5;  DB 1;  Length 513;
  Best Local Similarity   20.6%;  Pred. No. 0.042;
  Matches   90;  Conservative   70;  Mismatches  155;  Indels  121;  Gaps   18;

Qy         18 AAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRAGQHRGSATSLVFKQ 77
              |:  |  |    |:       :  |  ||::|     ||      |          | :|     :
Db        146 ASEESAPQGQDSAQDTTSESRELDNED-----RVDSKPEG----------GDSTQ---ES 187

Qy         78 KTITSW----------------MDTKGIKTVESESLHSKENNNTREESMMSSVQK-DNFY 120
              ::    |                :| :|::: :  ||: |:  |:    :  |  |  :  :|
Db        188 ESEEHWVGGGSDGESSHGDGSELDDEGMQSDDPESIRSERGNSRMNSAGMKSKESGENSE 247

Qy        121 QHNMEK------LENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPP 174
              | |  :        ||: |:  | || : :                  |   ||   |:
Db        248 QANTQDSGGSQLLEHPSRKIFRKSRISE------------------EDDRSE--LDDNNT 287

Qy        175 AVTLVPEQFSNAN---VDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGRE 231
                 :  :    |:|       | |: |      :  |:|::|                 : |: |  :
Db        288 MEEVKSDSTENSNSRDTGLSQPRRDSKGDSQEDSKEN-------------LSQEESQNVD 334

Qy        232 ARSHQKCGKACHPAEACAGCQQEE-------------TDVVSESPLSDTGSEDVGTGLKN 278
                 | :    :|    :: :   |||                | |  :   ||:  ||     | :
Db        335 GPSSESSQEANLSSQENSSESQEEVVSESRGDNPDPTTSYVEDQEDSDSSEEDSSHTLSH 394

Qy        279 ANRLNRQE---SSLGNSPPFEKESEPESPMDVDNSKN---SCQDSEADEETSPGFDEQED 332
              :   :|:|     |     |   |  :|| ||||  | ::|            | |: ::        |::|
Db        395 SKSESREEQADSESSESLNFSEES-PESPEDENSSSQEGLQSHSSSAESQSEESHSEEDD 453

Qy        333 SSSAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKRPGSTS 392
              | |   :        || :         ||   |  || :  |::: : :  |                     :
Db        454 SDSQDS----SRSKEDSNSTE-SKSSSEEDGQLK-NIEIE------------------SR 489

Qy        393 SLNVECRNSKQHGRKD 408
               | |:  ::|   | :|
Db        490 KLTVDAYHNKPIGDQD 505

RESULT 7
UN89_CAEEL
ID   UN89_CAEEL     STANDARD;      PRT;  6632 AA.
AC   O01761; Q17362;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN   UNC-89 OR C09D1.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=Bristol N2;
RX   MEDLINE=96180278; PubMed=8603916;
RA   Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT   "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT   assembly, encodes a giant modular protein composed of Ig and signal
RT   transduction domains.";
RL   J. Cell Biol. 132:835-848(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z., Le T.T., Wilson R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC       lattice assembly begins with positional cues laid down in the
CC       basement membrane and muscle cell membrane. UNC-89 responds to
CC       these signals, localizes, and then participates in assembling an
CC       M-line.
CC   -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC   -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 5 RCSD domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33058; AAB00542.1; -.
DR   EMBL; AF003131; AAB54132.2; -.
DR   PDB; 1FHO; 20-DEC-00.
DR   WormPep; C09D1.1; CE30426.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR007850; RCSD.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 47.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF05177; RCSD; 5.
DR   Pfam; PF00621; RhoGEF; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 23.
DR   SMART; SM00325; RhoGEF; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW   3D-structure.
FT   DOMAIN       63    127       SH3.
FT   DOMAIN      152    330       DH.
FT   DOMAIN      342    498       PH.
FT   DOMAIN      547    633       IG-LIKE C2-TYPE 1.
FT   DOMAIN      648    736       IG-LIKE C2-TYPE 2.
```

```
FT   DOMAIN      748    838       IG-LIKE C2-TYPE 3.
FT   DOMAIN      946   1033       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1044   1132       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1140   1227       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1272   1315       THR-RICH.
FT   DOMAIN     1375   1475       RCSD 1.
FT   DOMAIN     1479   1585       RCSD 2.
FT   DOMAIN     1597   1695       RCSD 3.
FT   DOMAIN     1700   1799       RCSD 4.
FT   DOMAIN     1800   1860       RCSD 5.
FT   DOMAIN     1982   2067       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2071   2163       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2171   2261       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2269   2359       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2367   2455       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2463   2564       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2563   2651       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2657   2746       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2754   2858       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2887   2980       IG-LIKE C2-TYPE 16.
FT   DOMAIN     2994   3081       IG-LIKE C2-TYPE 17.
FT   DOMAIN     3087   3183       IG-LIKE C2-TYPE 18.
FT   DOMAIN     3189   3280       IG-LIKE C2-TYPE 19.
FT   DOMAIN     3286   3376       IG-LIKE C2-TYPE 20.
FT   DOMAIN     3384   3469       IG-LIKE C2-TYPE 21.
FT   DOMAIN     3482   3572       IG-LIKE C2-TYPE 22.
FT   DOMAIN     3580   3667       IG-LIKE C2-TYPE 23.
FT   DOMAIN     3686   3777       IG-LIKE C2-TYPE 24.
FT   DOMAIN     3817   3908       IG-LIKE C2-TYPE 25.
FT   DOMAIN     3920   4009       IG-LIKE C2-TYPE 26.
FT   DOMAIN     4018   4106       IG-LIKE C2-TYPE 27.
FT   DOMAIN     4109   4201       IG-LIKE C2-TYPE 28.
FT   DOMAIN     4212   4297       IG-LIKE C2-TYPE 29.
FT   DOMAIN     4302   4387       IG-LIKE C2-TYPE 30.
FT   DOMAIN     4400   4485       IG-LIKE C2-TYPE 31.
FT   DOMAIN     4489   4580       IG-LIKE C2-TYPE 32.
FT   DOMAIN     4588   4678       IG-LIKE C2-TYPE 33.
FT   DOMAIN     4681   4771       IG-LIKE C2-TYPE 34.
FT   DOMAIN     4873   4961       IG-LIKE C2-TYPE 35.
FT   DOMAIN     4965   5057       IG-LIKE C2-TYPE 36.
FT   DOMAIN     5067   5160       IG-LIKE C2-TYPE 37.
FT   DOMAIN     5171   5260       IG-LIKE C2-TYPE 38.
FT   DOMAIN     5277   5366       IG-LIKE C2-TYPE 39.
FT   DOMAIN     5383   5472       IG-LIKE C2-TYPE 40.
FT   DOMAIN     5487   5578       IG-LIKE C2-TYPE 41.
FT   DOMAIN     5595   5685       IG-LIKE C2-TYPE 42.
FT   DOMAIN     5701   5790       IG-LIKE C2-TYPE 43.
FT   DOMAIN     5815   5904       IG-LIKE C2-TYPE 44.
FT   DOMAIN     5925   6014       IG-LIKE C2-TYPE 45.
FT   DOMAIN     6038   6130       IG-LIKE C2-TYPE 46.
FT   DOMAIN     6150   6239       IG-LIKE C2-TYPE 47.
FT   DOMAIN     6275   6368       FIBRONECTIN TYPE-III.
FT   DOMAIN     6413   6502       IG-LIKE C2-TYPE 48.
FT   DOMAIN     6507   6596       IG-LIKE C2-TYPE 49.
FT   DISULFID    568    621       POTENTIAL.
FT   DISULFID   2908   2975       POTENTIAL.
FT   DISULFID   3015   3065       POTENTIAL.
FT   DISULFID   3707   3759       POTENTIAL.
FT   DISULFID   3826   3890       POTENTIAL.
FT   DISULFID   5092   5157       POTENTIAL.
FT   DISULFID   5298   5350       POTENTIAL.
FT   DISULFID   5508   5560       POTENTIAL.
FT   DISULFID   5616   5669       POTENTIAL.
FT   DISULFID   5722   5764       POTENTIAL.
FT   DISULFID   5836   5901       POTENTIAL.
FT   DISULFID   5946   5998       POTENTIAL.
FT   DISULFID   6036   6171       POTENTIAL.
FT   DISULFID   6421   6486       POTENTIAL.
FT   CONFLICT   2137   2137       A -> P (IN REF. 1).
FT   CONFLICT   2245   2247       AKA -> PKP (IN REF. 1).
FT   CONFLICT   2258   2258       A -> P (IN REF. 1).
FT   CONFLICT   2284   2284       E -> G (IN REF. 1).
FT   CONFLICT   2297   2297       M -> I (IN REF. 1).
FT   CONFLICT   3531   3531       A -> G (IN REF. 1).
FT   CONFLICT   3884   3888       DAGEY -> RRRRI (IN REF. 1).
FT   CONFLICT   3929   3929       A -> V (IN REF. 1).
FT   CONFLICT   5134   5134       A -> P (IN REF. 1).
FT   CONFLICT   5145   5145       T -> S (IN REF. 1).
FT   CONFLICT   5185   5185       G -> A (IN REF. 1).
FT   CONFLICT   5199   5199       K -> N (IN REF. 1).
FT   CONFLICT   5202   5202       L -> F (IN REF. 1).
FT   CONFLICT   5213   5213       F -> L (IN REF. 1).
FT   CONFLICT   6178   6178       A -> G (IN REF. 1).
FT   CONFLICT   6268   6268       K -> E (IN REF. 1).
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;

  Query Match             3.1%;  Score 159;  DB 1;  Length 6632;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches  185;  Conservative  136;  Mismatches  387;  Indels  206;  Gaps   45;

Qy          9 PCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRAGQHRG 68

Db       1407 PKEKSPASPTKKEKSPAAEEVKS-PTKKEK---SPSSPTKKEKSPSSPTKKTGDEVKEKS 1462

Qy         69 SATSLVFKQKTITSWMDTK------------GIKTVESE-SLHSKENNNTRE-----ESM 110

Db       1463 PPKSPTKKEKSPEKPEDVKSPVKKEKSPDATNIVEVSSETTIEKTETTMTTEMTHESEES 1522

Qy        111 MSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLE 170

Db       1523 RTSVKKEKTPEKVDEKPKSPTKK--DKSPEKSITEEIKSPVK------KEKSPEKVEEKP 1574

Qy        171 SEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTME-DEQG 229

Db       1575 ASPTKKEKSPEK--PASPTKKSENEVKSPTKKEKSPE-KSVVEELKSPKEKSPEKADDKP 1631

Qy        230 REARSHQKCGKACHPAEACAGCQQEET-DVVSESPLSDTGSEDVGT-----GLKNANRLN 283

Db       1632 KSPTKKEKSPEKSATEDVKSPTKKEKSPEKVEEKPTSPTKKESSPTKKTDDEVKSPTKKE 1691

Qy        284 RQESSLGNSP--PFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANK 341

Db       1692 KSPQTVEEKPASPTKKEKSPE--------KSVVEEVKSPKEKSPEKAEEKPKSPTKKEKS 1743

Qy        342 PSRFQPREADTELRK-RSSAKGGEIRLHFQFEGGESRAGMNDVNAKRP----GSTSSLNV 396

Db       1744 PEKSAAEEVKSPTKKEKSPEKSAEEKPKSPTKKESSPVKMADDEVKSPTKKEKSPEKVEE 1803

Qy        397 ECRNSKQHGRKDSKITDHFMRVP------------KAEDKRKEQC-EMKHQRTERKIPKY 443

Db       1804 KPASPTKKEKTPEKSAAEELKSPTKKEKSPSSPTKKTGDESKEKSPEKPEEKPKSPTPKK 1863

Qy        444 IPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHF 503

Db       1864 SPPG-SPKKKKSKSP--EAEKPP-----APKLTRDLKLQTVNKTDLAHF-EVVVEHATEC 1914

Qy        504 KDLWDNKHVKMPCSEQNLYPVEDENGE-------------------RAAGSRWELIQTA 543

Db       1915 KWFLDGKEI---TTAQGVTVSKDDQFEFRCSIDTTMFGSGTVSVVASNAAGS----VETK 1967

Qy        544 L-LNRLTRPQNLKDAILKYNVAYSKKWDFT-ALIDF----WDKVLEEAEAQH--LYQ--- 592

Db       1968 TELKVLETPKE------------TKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQ 2015

Qy        593 ------------SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFF 640

Db       2016 NGNLLEDGKNGVTIKNEENKSSLIIPN--AQDSGKITVEASNEVG-SSESSAQLTVNPPS 2072

Qy        641 CT-----FPR-------RNAKMKSEYSSYP------DINFNRLFEGRSSRKPEKLKTLFC 682

Db       2073 TTPIVVDGPKSVTIKETETAEFKATISGFPAPTVKWTIN-EKIVE--ESRTITTIKTEDV 2129

Qy        683 YFRRVTEKK--PTGLVTFTRQSL--EDFPEWE-------RCEKLLTRL--HVTYEG---- 725

Db       2130 YTLKISNAKIEQTGTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLR 2189
```

```
Qy        726 ---TIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE-C 781
                 ::|   |                     |: :|:| : :      :|:|| :
Db       2190 WNLELDGPSPG----------------------TEVSWLLNGQPLTKSDTVQVVDHGDGT 2227

Qy        782 LIITGTEQYSEYTG 795
                 :|  |    | :|
Db       2228 YHVTIAEAKPEMSG 2241


RESULT 8
MDN1_HUMAN
ID   MDN1_HUMAN     STANDARD;      PRT;  5596 AA.
AC   Q9NU22; O15019;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Midasin (MIDAS-containing protein).
GN   MDN1 OR KIAA0301.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   PubMed=12102729;
RA   Garbarino J.E., Gibbons I.R.;
RT   "Expression and genomic analysis of midasin, a novel and highly
RT   conserved AAA protein distantly related to dynein.";
RL   BMC Genomics 3:18-18(2002).
RN   [2]
RP   SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.
RC   TISSUE=Testis;
RA   Tracey A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3550-5596 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC       the assembly/disassembly of macromolecular complexes in the
CC       nucleus.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 1 VWFA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF503925; AAM77722.1; -.
DR   EMBL; AL096678; CAB86660.1; -.
DR   EMBL; AL096678; CAB86661.1; -.
DR   EMBL; AB002299; BAA20761.1; -.
DR   Genew; HGNC:18302; MDN1.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0016887; F:ATPase activity; NAS.
DR   GO; GO:0003754; F:chaperone activity; NAS.
DR   GO; GO:0006461; P:protein complex assembly; NAS.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 7.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Chaperone; ATP-binding; Repeat; Nuclear protein.
FT   NP_BIND     329    336       ATP (POTENTIAL).
FT   NP_BIND     677    684       ATP (POTENTIAL).
FT   NP_BIND    1084   1091       ATP (POTENTIAL).
FT   NP_BIND    1390   1397       ATP (POTENTIAL).
FT   NP_BIND    1753   1760       ATP (POTENTIAL).
FT   NP_BIND    2066   2073       ATP (POTENTIAL).
FT   DOMAIN     3566   3573       POLY-GLU.
FT   DOMAIN     4784   4791       POLY-GLU.
FT   DOMAIN     5008   5013       POLY-GLU.
FT   DOMAIN     5182   5187       POLY-GLU.
FT   DOMAIN     5384   5583       VWFA.
FT   CONFLICT   2287   2312       RLFLSMDPVHGDISRAMRNRGLEIYI -> S (IN REF.
FT                                2).
SQ   SEQUENCE   5596 AA;  632802 MW;  586C62616A1F96D4 CRC64;

  Query Match             3.1%;  Score 158.5;  DB 1;  Length 5596;
  Best Local Similarity   18.5%;  Pred. No. 1;
  Matches  174;  Conservative  142;  Mismatches  312;  Indels  313;  Gaps   45;

Qy         90 KTVESESLHSKENNNTREESMMSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQYLKQ 149
              |  | |:|   :: |     |           ||     |: |  :  |   : | | :  ::
Db       4872 KVPEPEALDLPDDLNLDSEDKNGGEDTDN------EEGEEENPLEIKEKPEEAGHEAEER 4925

Qy        150 HQTAAMCKWQN---------EGPHSERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDT 200
               :|     ||         ||| :  |  |        |:     ||     | :
Db       4926 GETETD---QNESQSPQEPEEGPSEDDKAEGE--------EEMDTGADDQDGDAAQHPEE 4974

Qy        201 NSEESRDNQQFLTHVKLANAK-----------QTMEDEQGREARSHQKCGKACHPAEACA 249
              :||| :  |        | |: :           |  |:|:  :: : ::  :|   |  |
Db       4975 HSEEQQ--QSVEEKDKEADEEGGENGPADQGFQPQEEEEREDSDTEEQVPEALERKEH-A 5031

Qy        250 GCQQ---------EETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSL----------- 289
               | |          :  ::    :|  : | |: |:|  :||:     ||:
Db       5032 SCGQTGVENMQNTQAMELAGAAPEKEQGKEEHGSGAADANQAEGHESNFIAQLASQKHTR 5091

Qy        290 GNSPPFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPRE 349
              |:  |:::        |     ||  ::   :|    :    || |  :|  |   :  :  | :
Db       5092 KNTQSFKRK-----PGQADNERSMGDHNERVHKRLRTVD--TDSHAEQGPAQQPQAQVED 5144

Qy        350 ADT--ELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRK 407
              ||     ::: | |                  |    || :|                :|    |
Db       5145 ADAFEHIKQGSDAYD---------------AQTYDVASK--------------EQQQSAK 5175

Qy        408 DSKITDHFMRVPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPR 467
              ||              :|: :|: |      || :               :::    :|:::
Db       5176 DS-----------GKDQEEEEIEDTLMDTEEQ-----------EEFKAADVEQLK---- 5208

Qy        468 CGIRLPPLRPSANHTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDE 527
                    |     :   |   :   |  : : |:        :|            |  :::      |:|
Db       5209 ------PEEIKSGTTAPLGFDEMEV-EIQTVKTEEDQD----------PRTDKAHKETENE 5252

Qy        528 NGERAAGSRWELIQTA---LLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWD----- 579
                ||   ||    |   ||  |  ||    |:: : :|  |||         : :      : : |
Db       5253 KPER---SRESTIHTAHQFLMDTIFQP-FLKD-------VNELRQELERQLEMWQPRESG 5301

Qy        580 KVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAF 639
                 ||   |    ::|| |                |  |   |    | |::    :   :   |
Db       5302 NPEEEKVAAEMWQSYL------------ILTAP---LSQRLCEELRLILE---------- 5336

Qy        640 FCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSR-----------KPEKLKTLFCYFRRVT 688
                   | : ||:| :| :    :|   ::     :|:               || |  :   |
Db       5337 ----PTQAAKLKGDYRTGKRLNIRKVIPYIASQFRKDKIWLRRTKPSKRQYQICL---AI 5389

Qy        689 EKKPTGLVTFTRQSLEDFPEWERCEKLLTRLHVTYEG-TIEGNGQGMLQVDFANRFVGGG 747
              :    :  :      |:|                      : :|     :  ||    :|:|      |
Db       5390 DDSSSMVDNHTKQ-------------------LAFESLAVIGNALTLLEV---------G 5421

Qy        748 VTSAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHE 807
              :         |  :: |::|                       ||:|:|:|  ::  |   :
Db       5422 QIAVCSFGESVK-LLHP---------------------FHEQFSDYSG-SQILRLCK--- 5455
```

```
Qy        808 DRSERDDWQRRTTEIVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVA 867
                     :|::  |:|              ||: | :       |     |    :|||     :
Db       5456 -------FQQKKTKIA-----------QFL-ESVANMF--AAAQQLSQNISSETAQLLL 5493

Qy        868 TGNWGCGAFGGDARLKALIQILAAAVAERDV---VYFTFGDSELMRDIYSMHTFLTERKL 924
                 : | | |       |:    ::|||  | |:    | |    |:   ||        : : |:
Db       5494 VVSDGRGLF-----LEGKERVLAAVQAARNANIFVIFVVLDNPSSRD------SILDIKV 5542

Qy        925 TV----GEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 961
                :      ||: :  :| | ||                  :|| |: :
Db       5543 PIFKGPGEMPE--IRSYMEE-------------FPPPYYII 5568
```

```
RESULT 9
ELA1_MOUSE
ID   ELA1_MOUSE     STANDARD;      PRT;   773 AA.
AC   Q8CB77; Q80VB2; Q9R0Q5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcription elongation factor B polypeptide 3 (RNA polymerase II
DE   transcription factor SIII subunit A) (SIII p110) (Elongin A) (EloA)
DE   (Elongin 110 kDa subunit).
GN   TCEB3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20044622; PubMed=10575222;
RA   Aso T., Amimoto K., Takebayashi S., Okumura K., Hatakeyama M.;
RT   "Structural organization and chromosome location of the mouse elongin
RT   A gene (Tceb3).";
RL   Cytogenet. Cell Genet. 86:259-262(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Bone;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: SIII, also known as elongin, is a general transcription
CC       elongation factor that increases the RNA polymerase II
CC       transcription elongation past template-encoded arresting sites.
CC       Subunit A is transcriptionally active and its transcription
CC       activity is strongly enhanced by binding to the dimeric complex of
CC       the SIII regulatory subunits B and C (elongin BC complex) (By
CC       similarity).
CC   -!- SUBUNIT: Heterotrimer of an A (A1, A2 or A3), B and C subunit (By
CC       similarity).
CC   -!- DOMAIN: The elongin BC complex binding domain is also known as BC-
CC       box with the consensus [APST]-L-x(3)-C-x(3)-[AILV].
CC   -!- SIMILARITY: Contains 1 F-box domain.
CC   -!- SIMILARITY: Contains 1 TFS2-N domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB025015; BAA84994.1; -.
DR   EMBL; AK036592; BAC29497.1; -.
DR   EMBL; BC049885; AAH49885.1; -.
DR   MGD; MGI:1351315; Tceb3.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR003617; TFS2_N.
DR   SMART; SM00509; TFS2N; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Transcription regulation; Nuclear protein.
FT   DOMAIN        1     78       TFS2-N.
FT   DOMAIN      521    680       ACTIVATION DOMAIN (BY SIMILARITY).
FT   DOMAIN      549    558       INTERACTING WITH ELONGIN BC COMPLEX (BY
FT                                SIMILARITY).
FT   DOMAIN      565    609       F-BOX.
FT   CONFLICT     35     35       I -> V (IN REF. 1).
FT   CONFLICT    142    142       P -> R (IN REF. 1).
FT   CONFLICT    159    159       H -> Q (IN REF. 1).
FT   CONFLICT    236    236       H -> Q (IN REF. 1).
FT   CONFLICT    382    382       L -> P (IN REF. 2).
FT   CONFLICT    609    609       H -> P (IN REF. 1).
FT   CONFLICT    625    626       WR -> LE (IN REF. 1).
FT   CONFLICT    633    635       QDA -> SER (IN REF. 1).
SQ   SEQUENCE   773 AA;  87177 MW;  F6DC1E821B3DB5FF CRC64;

  Query Match             3.0%;  Score 156.5;  DB 1;  Length 773;
  Best Local Similarity   20.6%;  Pred. No. 0.11;
  Matches  124;  Conservative   78;  Mismatches  204;  Indels  195;  Gaps   31;

Qy         79 TITSWMDTKGIKTVESESLHSKENNNTRE---------------ESMMSSVQKDNFYQHN 123
              |:   ::|    ||| |    | :  |  |:                |:      :|:|   :
Db         39 TVDILVETGVGKTVNSFRKHEQVGNFARDLVAQWKKLVPVERNSEAEDQDFEKNNSRKRP 98
```

```
Qy        124 MEKLENVSQL--GFDKSPVEKGTQ-YLKQHQTAAMCKW-QNEGPHS-----------ERL 168
                :  |:    :|    : :|     |:: |  | :|:        |  : | ||              :|
Db         99 RDALQREEELEGNYQESWKPSGSRSYSPEHRQKKHKKLSEPERPHKVAHSHEKRDERKRC 158

Qy        169 LESEPPAVTLVPEQFSNANVDQ---SSPKDDHSDTNSEESRDNQQFLTHVKLANA-KQTM 224
                :  ||   :  ||         :|          |||    ::|  :        |  :  ::|  |          |
Db        159 HKVSPP-YSSDPESSDYGHVQSPPPSSPHQMYTDLSRSPEEDQEPIISHQKPGKVHSNTF 217

Qy        225 ED--------EQGREARSHQK----CGKACHPAEACAGCQQEETDVVSESPLSDTGSEDV 272
              :|           |||:  | ||  |          |  |||:|     :  |:     ||
Db        218 QDRLGVSHLGEQGKGAVSHHKQHRSSHKEKHPADA---REDEKISAVSRE---------- 264

Qy        273 GTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKNSCQDSEADEETSPGFDEQED 332
                    |:    :::||     |       ||  | |  :    |    :   :    |    ::  ||   |    |
Db        265 ----KSHKASSKEESRRLLSGDSAKEKLPSSVVKKDKDR---EGSSLKKKFSPALDVASD 317

Qy        333 SSSAQTANKPSRFQPREADTELRKRS-----SAKG---------GEIRLHFQFEGGESR- 377
                :          ||        :|  ::   |:|        | :|               ::   |  :  :  |:  |
Db        318 NH----FKKPKHKDSEKAKSDKNKQSVDGVDSGRGTGDPLPKAKEKVPNHLKAQEGKVRT 373

Qy        378 ------AGM-------NDVN---------------------AKRPGSTSSLNVECRNSK 402
                          ||:         ||:                                |:    |||   :    :    |
Db        374 NADGKSAGLLHPKAEETDVDDEFERPTMSFESYLSYDQPRKKKKKVVKTSSTALGEKGLK 433

Qy        403 QHGRKDSKIT----DHFMRVPKAEDKRKEQCEM--KHQRTERKIP--------------- 441
                  :||||  |        :      ::||     :  :  |:  :                     ||:|
Db        434 ---KKDSKSTSKNLNSAQKLPKVNENKSEKLQPAGAEPTRPRKVPTDVLPALPDIPLPAI 490

Qy        442 --KYIP-------PHLSPDKKWLGTPIEEM------RRM---------PRCGIRLPPLRP 477
                    |  |              |      |  :|      :|  ||             |||                 :|       ||  :
Db        491 HANYRPLPSLELIPSFQPKRKAFSSPQEEEEAGFTGRRMNSKMQVYSGSKCAY-LPKM-- 547

Qy        478 SANHTVTIRV------DLLRIGEVP----KPFPTHFKDLWDNKHVKMPCSEQNLYPVEDE 527
                    |     |||               :    :|   ||         :|                  |        |:      ||   :|:
Db        548 MTLHQQCIRVLKNNIDSIFEVGGVPYSVLEP------------VLERCTPDQLYRIEEC 594

Qy        528 N 528
              |
Db        595 N 595
```

```
RESULT 10
MINT_MOUSE
ID   MINT_MOUSE     STANDARD;      PRT;  3644 AA.
AC   Q62504; Q80TN9; Q99PS4; Q9QZW2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Msx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN   MINT OR SHARP OR KIAA0929.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
RP   LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC   TISSUE=Testis;
RX   MEDLINE=99379811; PubMed=10451362;
RA   Newberry E.P., Latifi T., Towler D.A.;
RT   "The RRM domain of MINT, a novel msx2 binding protein, recognizes and
RT   regulates the rat osteocalcin promoter.";
RL   Biochemistry 38:10678-10690(1999).
RN   [2]
RP   SEQUENCE OF 1-112 FROM N.A.
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA   Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA   Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA   Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA   Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA   Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA   Watahiki A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP   PHE-762; PHE-773 AND LEU-933.
RC   STRAIN=ICR; TISSUE=Brain;
RA   Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT   "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT   J/Su(H) activity.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 318-578 FROM N.A.
RC   TISSUE=Cochlea;
RX   MEDLINE=97237053; PubMed=9119401;
RA   Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA   Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA   Weil D., Pujol R., Petit C.;
RT   "Cloning of the genes encoding two murine and human cochlear
RT   unconventional type I myosins.";
RL   Genomics 40:332-341(1997).
RN   [5]
RP   SEQUENCE OF 2598-3644 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
RN   [6]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=22261914; PubMed=12374742;
RA   Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA   Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA   Schmid R.M.;
RT   "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT   pathway.";
RL   EMBO J. 21:5417-5426(2002).
RN   [7]
RP   FUNCTION, AND TISSUE SPECIFICITY.
RX   MEDLINE=22483652; PubMed=12594956;
RA   Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA   Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT   "Regulation of marginal zone B cell development by MINT, a suppressor
RT   of Notch/RBP-J signaling pathway.";
RL   Immunity 18:301-312(2003).
CC   -!- FUNCTION: Essential corepressor protein, which probably regulates
CC       different key pathways such as the Notch pathway. Negative
CC       regulator of the Notch pathway via its interaction with RBPSUH,
CC       which prevents the association between NOTCH1 and RBPSUH, and
CC       therefore suppresses the transactivation activity of Notch
CC       signaling. Blocks the differentiation of precursor B cells into
CC       marginal zone B cells. Probably represses transcription via the
CC       recruitment of large complexes containing histone deacetylase
CC       proteins. May bind both to DNA and RNA.
CC   -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and
CC       MTA1L1. Interacts with the nuclear receptors RAR and PPARD.
CC       Interacts with RAR in absence of ligand. Bind to the steroid
CC       receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
CC       Interacts with RBPSUH; this interaction may prevent the
CC       interaction between RBPSUH and NOTCH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q62504-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q62504-2; Sequence=VSP_008564;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower
CC       level in brain, lung, spleen, liver and kidney. Weakly expressed
CC       in cardiac and skeletal muscles and ovary. In spleen, it is
```

```
CC       expressed in follicular B-cells, while it is weakly expressed in
CC       marginal zone B-cells.
CC   -!- DOMAIN: The RID domain mediates the interaction with nuclear
CC       receptors.
CC   -!- DOMAIN: The SPOC domain, which mediates the interaction with
CC       NCOR2, is essential for the repressive activity (By similarity).
CC   -!- SIMILARITY: Belongs to the Spen family.
CC   -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC   -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC   -!- SIMILARITY: Contains 1 SPOC domain.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to multiple
CC       frameshifts and conflicts that create stop codons.
CC   -!- CAUTION: Ref.5 sequence differs from that shown due to what seems
CC       to be the presence of intronic sequence in the cDNA.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; BY726481; -; NOT_ANNOTATED_CDS.
DR   EMBL; AF156529; AAD55931.1; ALT_INIT.
DR   EMBL; AB055980; BAB32786.1; -.
DR   EMBL; Z78160; CAB01562.1; ALT_SEQ.
DR   EMBL; AK122402; BAC65684.2; ALT_SEQ.
DR   MGD; MGI:1891706; Mint.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 3.
DR   SMART; SM00360; RRM; 3.
DR   PROSITE; PS50102; RRM; 4.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR   PROSITE; PS50917; SPOC; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding;
KW   RNA-binding; Repeat; Coiled coil; Alternative splicing; Polymorphism.
FT   DOMAIN        1    574       DNA-BINDING.
FT   DOMAIN        6     81       RNA-BINDING (RRM) 1.
FT   DOMAIN      336    416       RNA-BINDING (RRM) 2.
FT   DOMAIN      439    514       RNA-BINDING (RRM) 3.
FT   DOMAIN      518    590       RNA-BINDING (RRM) 4.
FT   DOMAIN      559    575       COILED COIL (POTENTIAL).
FT   DOMAIN      822    850       COILED COIL (POTENTIAL).
FT   DOMAIN     1185   1206       COILED COIL (POTENTIAL).
FT   DOMAIN     1509   1544       COILED COIL (POTENTIAL).
FT   DOMAIN     1607   1627       COILED COIL (POTENTIAL).
FT   DOMAIN     2216   2704       RID.
FT   DOMAIN     3478   3644       SPOC.
FT   DOMAIN     2138   2462       INTERACTION WITH MSX2.
FT   DOMAIN     2706   2845       INTERACTION WITH RBPSUH.
FT   DOMAIN      125    277       ARG-RICH.
FT   DOMAIN      236    326       SER-RICH.
FT   DOMAIN      648    721       TYR-RICH.
FT   DOMAIN      702    832       ARG-RICH.
FT   DOMAIN     2101   2233       ALA-RICH.
FT   DOMAIN     2377   2518       PRO-RICH.
FT   DOMAIN     2950   3475       PRO-RICH.
FT   VARSPLIC    618    640       Missing (in isoform 2).
FT                                /FTId=VSP_008564.
FT   VARIANT     348    348       I -> T.
FT   VARIANT     762    762       S -> F.
FT   VARIANT     773    773       S -> F.
FT   VARIANT     933    933       S -> L.
FT   CONFLICT    754    754       R -> G (IN REF. 3).
FT   CONFLICT   1524   1524       D -> A (IN REF. 3).
FT   CONFLICT   1560   1560       H -> Y (IN REF. 3).
FT   CONFLICT   1570   1570       F -> L (IN REF. 3).
FT   CONFLICT   1574   1574       R -> G (IN REF. 3).
FT   CONFLICT   1609   1609       Q -> R (IN REF. 3).
FT   CONFLICT   1659   1659       I -> V (IN REF. 3).
FT   CONFLICT   1669   1669       S -> F (IN REF. 3).
FT   CONFLICT   1705   1705       V -> A (IN REF. 3).
FT   CONFLICT   1815   1815       A -> V (IN REF. 3).
FT   CONFLICT   2097   2097       G -> A (IN REF. 3).
FT   CONFLICT   2201   2202       MISSING (IN REF. 3).
FT   CONFLICT   2322   2322       A -> V (IN REF. 3).
FT   CONFLICT   2385   2385       P -> Q (IN REF. 3).
FT   CONFLICT   2502   2502       R -> K (IN REF. 3).
FT   CONFLICT   2505   2505       E -> K (IN REF. 3).
FT   CONFLICT   2519   2519       D -> N (IN REF. 3).
FT   CONFLICT   2554   2554       T -> S (IN REF. 3).
FT   CONFLICT   2679   2688       LVSTPAGPVN -> VGEHPWWARD (IN REF. 3).
FT   CONFLICT   3010   3010       L -> P (IN REF. 3 AND 5).
SQ   SEQUENCE   3644 AA;  398750 MW;  9C7EC49A81A7DA4A CRC64;

  Query Match             3.0%;  Score 156;  DB 1;  Length 3644;
  Best Local Similarity   19.2%;  Pred. No. 0.83;
  Matches  128;  Conservative   86;  Mismatches  203;  Indels  250;  Gaps   31;

Qy         15 RWDAAATSPPAASDARSFPG--------RQRRVLDSKDAPVQFRVPPSSSGCALGRAGQH 66
              |||        |:  |  |||          |:| | |  :    |     || |       :|
Db       1354 RWDPPMKQDPSRFDV-SFPNSVIKRDSLRKRSVRDLEPGEV-----PSDS----DEDAEH 1403

Qy         67 RGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENN-NTREESMMSSVQKDNFYQHNME 125
              |       :     :|: |: :     |   |    :    |:| : ||:::: ||     ::
Db       1404 RS-------QSPRASSFYDSPRL----SFLLRDRDQKLRERDERLASSLERNKFYSFALD 1452

Qy        126 KL--------------------ENVSQLGF-----------DKSPVEKGTQ-----YLKQ 149
              |                     || | | :            ||  |:    :      |:|:
Db       1453 KTITPDTKALLERAKSLSSSREENWSFLDWDSRFANFRNNKDKEKVDSAPRPIPSWYMKK 1512

Qy        150 HQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANV-DQSSPKDDHSDTNSEES--- 205
                :    : : :       ||  ||      |    :| :::: :| |  :   |  :  ||||
Db       1513 KKIRTDSEGKLDDKKDER-REEEQERQELFASRFLHSSIFEQDSKRLQHLERKSEESDFP 1571

Qy        206 ----------------RDNQQ---FLTHVKL-----ANAKQTMEDEQGREARSHQKCGKA 241
                              |: |    | |  :            |:  :|:: :||    ::
Db       1572 PGRLYGRQASEGANSTSDSVQEPVVLFHSRFMELTRMQQKEKEKDQKPKEAEKQEE--PE 1629

Qy        242 CHP---------------AEACAGCQ---------------------------------- 252
              ||                 |    ||
Db       1630 THPKTPEPAAETKEPEPKAPVSAGLPAVTITVVTPEPASSAPEKAEEAAEAPSPAGEKPA 1689

Qy        253 -----QEETDVVSES---PLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEK------- 297
                   ||| :|||       |:          ||  |      :| ::  ::|  | | |
Db       1690 EPAPVSEETKLVSEPVSVPVEQPRQSDVPPG--EDSRDSQDSAALAPSAPQESAATDAVP 1747

Qy        298 --------------------ESEPESPMDVDNSKNSCQDSEADE------ETSPGFD--- 328
                                  : :|  ||   :    ||  :  :   ||:|         :|:|    :
Db       1748 CVNAEPLTPGTTVSQVESSVDPKPSSPQPL--SKLTQRSEEAEEGKVEKPDTTPSTEPDA 1805

Qy        329 -------------EQED-------------------------SSSAQTANKPSRFQPREA 350
                           ||                          :::|     ||     :
Db       1806 TQNAGVASEAQPPASEDVEANPPVAAKDRKTNKSKRSKTSVQAAAASVVEKPVTRKSERI 1865

Qy        351 DTELRKRSSAKGGEIR--LHFQFEGGE-----SRAGMNDVNAKRPGSTSSLNVECRNSKQ 403
              |  |   ||||:   || :    |     :  | |::    |         |        |: ::
Db       1866 DREKLKRSSSPRGEAQKLLELKMEAEKITRTASKSSGGDTEHPEPSLPLS-----RSRRR 1920

Qy        404 HGRK-DSKITDHFMRVPKAEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEM 462
              : |    : :|||  |  |   |      ||   :   :| ||:: : :  |    :| :    | |  :
Db       1921 NVRSVYATMTDHESRSPAKEP--VEQPRVTRKRLERELQEAVVPPTTPRR---GRPPKTR 1975

Qy        463 RRMPRCG 469
              ||    |
Db       1976 RRAEEDG 1982

RESULT 11
MINT_HUMAN
ID   MINT_HUMAN     STANDARD;      PRT;  3664 AA.
AC   Q96T58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
```

```
DE   Msx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN   MINT OR SHARP OR KIAA0929.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
RP   RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
RP   RAR AND MTA1L1.
RC   TISSUE=Liver, and Pituitary;
RX   MEDLINE=21231190; PubMed=11331609;
RA   Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
RA   Hon M., Evans R.M.;
RT   "Sharp, an inducible cofactor that integrates nuclear receptor
RT   repression and activation.";
RL   Genes Dev. 15:1140-1151(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bird C.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 294-3664 FROM N.A.
RA   Rhodes S., Huckle E.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC   TISSUE=Embryo, and Teratocarcinoma;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 2002-3664 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [6]
RP   INTERACTION WITH PPARD.
RX   MEDLINE=21874127; PubMed=11867749;
RA   Shi Y., Hon M., Evans R.M.;
RT   "The peroxisome proliferator-activated receptor delta, an integrator
RT   of transcriptional repression and nuclear receptor signaling.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
RN   [7]
RP   FUNCTION, AND INTERACTION WITH RBPSUH.
RX   MEDLINE=22261914; PubMed=12374742;
RA   Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA   Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA   Schmid R.M.;
RT   "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT   pathway.";
RL   EMBO J. 21:5417-5426(2002).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
RX   MEDLINE=22777836; PubMed=12897056;
RA   Ariyoshi M., Schwabe J.W.R.;
RT   "A conserved structural motif reveals the essential transcriptional
RT   repression function of Spen proteins and their role in developmental
RT   signaling.";
RL   Genes Dev. 17:1909-1920(2003).
CC   -!- FUNCTION: Essential corepressor protein, which probably regulates
CC       different key pathways such as the Notch pathway. Negative
CC       regulator of the Notch pathway via its interaction with RBPSUH,
CC       which prevents the association between NOTCH1 and RBPSUH, and
CC       therefore suppresses the transactivation activity of Notch
CC       signaling. Blocks the differentiation of precursor B cells into
CC       marginal zone B cells. Probably represses transcription via the
CC       recruitment of large complexes containing histone deacetylase
CC       proteins. May bind both to DNA and RNA.
CC   -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
CC       NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTA1L1. Interacts with
CC       RBPSUH; this interaction may prevent the interaction between
CC       RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and
CC       PPARD. Interacts with RAR in absence of ligand. Bind to the
CC       steroid receptor RNA coactivator SRA.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC   -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
CC       spleen and thymus. Expressed at intermediate level in kidney,
CC       liver, mammary gland and skin.
CC   -!- INDUCTION: By hormone 17-beta-estradiol (E2).
CC   -!- DOMAIN: The RID domain mediates the interaction with nuclear
CC       receptors (By similarity).
CC   -!- DOMAIN: The SPOC domain, which mediates the interaction with
CC       NCOR2, is essential for the repressive activity.
CC   -!- SIMILARITY: Belongs to the Spen family.
CC   -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC   -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC   -!- SIMILARITY: Contains 1 SPOC domain.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC       gene model prediction.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF356524; AAK52750.1; -.
DR   EMBL; AL034555; CAB85442.1; ALT_SEQ.
DR   EMBL; AL034555; CAB85444.1; ALT_SEQ.
DR   EMBL; AL450998; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL096858; CAB51072.1; ALT_INIT.
DR   EMBL; AK000882; BAA91405.1; ALT_INIT.
DR   EMBL; AK022949; BAB14324.1; ALT_INIT.
DR   EMBL; AB023146; BAA76773.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   PDB; 1OW1; 19-AUG-03.
DR   Pfam; PF00076; rrm; 4.
DR   SMART; SM00360; RRM; 4.
DR   PROSITE; PS50102; RRM; 4.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR   PROSITE; PS50917; SPOC; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding;
KW   Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
FT   DOMAIN        1    573       DNA-BINDING (BY SIMILARITY).
FT   DOMAIN        6     81       RNA-BINDING (RRM) 1.
FT   DOMAIN      335    415       RNA-BINDING (RRM) 2.
FT   DOMAIN      438    513       RNA-BINDING (RRM) 3.
FT   DOMAIN      517    589       RNA-BINDING (RRM) 4.
FT   DOMAIN      688    715       COILED COIL (POTENTIAL).
FT   DOMAIN      977   1004       COILED COIL (POTENTIAL).
FT   DOMAIN     1170   1191       COILED COIL (POTENTIAL).
FT   DOMAIN     1408   1428       COILED COIL (POTENTIAL).
FT   DOMAIN     1496   1529       COILED COIL (POTENTIAL).
FT   DOMAIN     1592   1612       COILED COIL (POTENTIAL).
FT   DOMAIN     1928   1944       COILED COIL (POTENTIAL).
FT   DOMAIN     2201   2707       RID.
FT   DOMAIN     3498   3664       SPOC.
FT   DOMAIN     2130   2464       INTERACTION WITH MSX2 (BY SIMILARITY).
FT   DOMAIN     2709   2870       INTERACTION WITH RBPSUH (BY SIMILARITY).
FT   DOMAIN      125    277       ARG-RICH.
FT   DOMAIN      240    325       SER-RICH.
FT   DOMAIN      616    810       ARG-RICH.
FT   DOMAIN      624    697       TYR-RICH.
FT   DOMAIN     2428   2520       PRO-RICH.
```

```
FT   DOMAIN     3220   3482       PRO-RICH.
FT   VARIANT     970    970       A -> V (in dbSNP:848208).
FT                                /FTId=VAR_017119.
FT   VARIANT    1091   1091       L -> P (in dbSNP:848209).
FT                                /FTId=VAR_017120.
FT   VARIANT    2360   2360       N -> D (in dbSNP:848210).
FT                                /FTId=VAR_017121.
FT   CONFLICT    956    956       G -> D (IN REF. 4).
SQ   SEQUENCE   3664 AA;  402245 MW;  5228C58533E5B27B CRC64;

  Query Match              3.0%;  Score 156;  DB 1;  Length 3664;
  Best Local Similarity   18.8%;  Pred. No. 0.84;
  Matches  101;  Conservative   78;  Mismatches  232;  Indels  126;  Gaps   19;

Qy          4 GPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCALGRA 63

Db       1994 GPQGKKGKNEPKVD--ATRPEATTEV----GPQIGVKESSMEP---KAAEEEAGSEQKRD 2044

Qy         64 GQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFYQHN 123

Db       2045 RKDAG----------TDKNPPETAPVEVVEKKPAPEKNSKSKRGRSRNSRLAVD-----K 2089

Qy        124 MEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSE----RLLESEPPAVTLV 179

Db       2090 SASLKNVDAAVSPRGAAAQAGE--RESGVVAVSPEKSESPQKEDGLSSQLKSDPVDPDKE 2147

Qy        180 PEQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHV-KLANAKQTMEDEQGREARSHQKC 238

Db       2148 PEK---EDVSASGPSPE--ATQLAKQMELEQAVEHIAKLAEAS----------------- 2185

Qy        239 GKACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKN-ANRLNRQESSLGNSPPFEK 297

Db       2186 ASAAYKADAPEGLAPEDRD----KPAHQASETELAAAIGSIINDISGEPENFPAPPPYPG 2241

Qy        298 ESE--------------PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPS 343

Db       2242 ESQTDLQPPAGAQALQPSEEGMETDEAVSGILETEAATESSRPPVNAPDPSAGPTDTKEA 2301

Qy        344 RFQPREADTELRKRSSAKGGEIRLHFQFEG---------------------------- 373

Db       2302 RGNSSETSHSVPEAKGSKEVEVTLVRKDKGRQKTTRSRRKRNTNKKVVAPVESHVPESNQ 2361

Qy        374 --GESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPKAEDKRKEQCEM 431

Db       2362 AQGESPAANEGTTVQHPEAPQE---EKQSEKPHSTPPQSCTSDLSKIPSTENSSQEISVE 2418

Qy        432 KHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRL----------PPLRPS 478

Db       2419 ERTPTKASVPPDLPPPPQP------APVDE---EPQARFRVHSIIESDPVTPPSDPS 2466

RESULT 12
DMP1_RAT
ID   DMP1_RAT       STANDARD;      PRT;   489 AA.
AC   P98193;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE   protein-1) (DMP-1) (AG1).
GN   DMP1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Tooth;
RX   MEDLINE=93286101; PubMed=8509401;
RA   George A., Sabsay B., Simonian P.A., Veis A.;
RT   "Characterization of a novel dentin matrix acidic phosphoprotein.
RT   Implications for induction of biomineralization.";
RL   J. Biol. Chem. 268:12624-12630(1993).
CC   -!- FUNCTION: Could be involved in the induction of mineralization of
CC       extracellular matrix.
CC   -!- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast
CC       and ameloblast.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L11354; -; NOT_ANNOTATED_CDS.
KW   Biomineralization; Extracellular matrix; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    489       DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT   DOMAIN       41     44       POLY-PRO.
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        334    336       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   489 AA;  53058 MW;  59F8381479DDA085 CRC64;

  Query Match              3.0%;  Score 155;  DB 1;  Length 489;
  Best Local Similarity   21.1%;  Pred. No. 0.071;
  Matches   78;  Conservative   71;  Mismatches  136;  Indels   84;  Gaps   17;

Qy         94 SESLHSKENNNTREESMMS--SVQKDNFYQHNMEKLENVSQLGFDKSPVEK-------GT 144

Db        122 ADTTQSSEDSTSQENSAQDTPSDSKD----HHSDEADSRPEAGDSTQDSESEEYRVGGGS 177

Qy        145 QYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPE---QFSNANVDQSSPKDDHSDTN 201

Db        178 EGESSHGDGS--EFDDEG------MQSDDPGSTRSDRGHTRMSSAGIRSEESKGDHEPTS 229

Qy        202 SEESRDNQ--QFLTHVKLANAKQTMEDEQG-------REARS------HQKCGKACHPAE 246

Db        230 TQDSDDSQDVEFSSRKSFRRSRVSEEDDRGELADSNSRETQSVSTEDFRSKEESRSETQE 289

Qy        247 ACAGCQQEETDVVSESPLSDTGSEDVG-----------TGLKNANRLNRQESSLGNSPPF 295

Db        290 DTAETQSQEDSPEGQDPSSES-SEEAGEPSQESSSESQEGVASESRGDNPDNTSQTGDQR 348

Qy        296 EKESEPESPMDVDNSKNSCQ-----DSEADEETSPGFDEQ-----EDSSS-----AQTAN 340

Db        349 DSESSEEDRLNTFSSSESQSTEEQGDSESNESLSLSEESQESAQDEDSSSQEGLQSQSAS 408

Qy        341 KPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSS------- 393

Db        409 RESRSQESQSEQDSRSEENRDS---------DSQDSSRSKEESNS--TGSTSSSEEDNHP 457

Qy        394 LNVECRNSK 402

Db        458 KNIEADNRK 466

RESULT 13
SR40_YEAST
ID   SR40_YEAST     STANDARD;      PRT;   406 AA.
AC   P32583;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Suppressor protein SRP40.
GN   SRP40 OR YKR092C OR YKR412A.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 28383 / FL100;
```

```
RA   Lalo D., Carles C., Sentenac A., Thuriaux P.;
RL   Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: Not known; weak suppressor of a mutant of the
CC       subunit AC40 of DNA dependant RNA polymerase I and III.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L11275; AAA35091.1; -.
DR   EMBL; X73541; CAA51946.1; -.
DR   EMBL; Z28317; CAA82171.1; -.
DR   PIR; S38170; S38170.
DR   GermOnline; 140071; -.
DR   SGD; S0001800; SRP40.
DR   GO; GO:0005730; C:nucleolus; IDA.
DR   InterPro; IPR007718; SRP40_C.
DR   Pfam; PF05022; SRP40_C; 1.
FT   DOMAIN       25    314       ASP/SER-RICH.
FT   CONFLICT    400    400       G -> N (IN REF. 1).
SQ   SEQUENCE   406 AA;  41015 MW;  8EA007695AF4BA1D CRC64;

  Query Match             3.0%;  Score 154;  DB 1;  Length 406;
  Best Local Similarity   18.6%;  Pred. No. 0.064;
  Matches   73;  Conservative   73;  Mismatches  174;  Indels   72;  Gaps   11;

Qy         40 LDSKDAPVQFRVPPSSSGCALGRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHS 99
              | |:  :: :     |||  :    :       |::|    : : :|    :     : |:| |
Db         14 LSVKEKEIEEKSSSSSSSSSSSSSSSSSSSSSSSSSGESSSSSSSSSSSSSSDSSDSSDS 73

Qy        100 KENNNTREESMMSSVQKDNFYQHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCK-W 158
              : ::::    |   ||    |:       :    : |        |   |  ::    : :|     :
Db         74 ESSSSSSSSSSSSSSSDSESSSESDSSSSGSSSSSSSSSDESSSESESEDETKKRARES 133

Qy        159 QNEGPHSERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLA 218
                ||      :   ::|| : :           |: :      |       |::  ||    |:                 :
Db        134 DNEDAKETKKAKTEPESSS------SSESSSSGSSSSSESESGSESDSDSS--------S 179

Qy        219 NAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKN 278
              ::  :   :  |      |: |                    :: :       ::      |:|    ||: |:
Db        180 SSSSSSDSESDSESDS-----------QSSSSSSSSDSSSDSDSSSSDSSSDS------- 221

Qy        279 ANRLNRQESSLGNSPPFEKESEPESPMDVDNS----KNSCQDSEADEETSPGFDEQEDSS 334
                          ||  :|     : :|: :|   | |:|          :|   ||  :|| ||         :   ||
Db        222 ------DSSSSSSSSSDSDSDSDSSSDSDSSGSSDSSSSSDSSSDESTS---SDSSDSD 272

Qy        335 SAQTANKPSRFQPREADTELRKRSSAKGGEIRLHFQFEGGESRAGMND--VNAKRPGSTS 392
              |     :     |   : :|| :   |                               |:  |    |:     :|      |  :
Db        273 SDSDSGSSSELETKEATADESKAE----------------ETPASSNESTPSASSSSSAN 316

Qy        393 SLNVECRNS--KQHGRKDSKITDHFMRVPKAE 422
                ||:          |:  ||            || ||  :::
Db        317 KLNIPAGTDEIKEGQRK------HFSRVDRSK 342


RESULT 14
IF38_HUMAN
ID   IF38_HUMAN     STANDARD;      PRT;   913 AA.
AC   Q99613; O00215;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Eukaryotic translation initiation factor 3 subunit 8 (eIF3 p110)
DE   (eIF3c).
GN   EIF3S8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97150873; PubMed=8995409;
RA   Asano K., Kinzy T.G., Merrick W.C., Hershey J.W.B.;
RT   "Conservation and diversity of eukaryotic translation initiation
RT   factor eIF3.";
RL   J. Biol. Chem. 272:1101-1109(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99425270; PubMed=10493829;
RA   Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA   Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA   Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA   Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT   "Genome duplications and other features in 12 Mb of DNA sequence from
RT   human chromosome 16p and 16q.";
RL   Genomics 60:295-308(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA   Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Binds to the 40S ribosome and promotes the binding of
CC       methionyl-tRNAi and mRNA.
CC   -!- SUBUNIT: eIF-3 is composed of at least 12 different subunits.
CC   -!- SIMILARITY: Contains 1 PCI domain.
CC   -!- SIMILARITY: BELONGS TO THE EIF3S8 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U46025; AAD03462.1; -.
DR   EMBL; AC002544; AAC27426.1; -.
```

```
DR   EMBL; U91326; AAC27674.1; -.
DR   EMBL; AK000739; BAA91352.1; -.
DR   EMBL; BC001571; AAH01571.1; -.
DR   Genew; HGNC:3279; EIF3S8.
DR   GK; Q99613; -.
DR   MIM; 603916; -.
DR   GO; GO:0005852; C:eukaryotic translation initiation factor 3 . . .; TAS.
DR   GO; GO:0003743; F:translation initiation factor activity; TAS.
DR   GO; GO:0006446; P:regulation of translational initiation; TAS.
DR   InterPro; IPR008905; eIF3c_N.
DR   InterPro; IPR000717; PCI.
DR   Pfam; PF05470; eIF3c_N; 1.
DR   Pfam; PF01399; PCI; 1.
DR   SMART; SM00088; PINT; 1.
KW   Initiation factor; Protein biosynthesis.
FT   DOMAIN      164    189       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      243    246       POLY-GLU.
FT   DOMAIN      291    294       POLY-GLU.
FT   CONFLICT    313    314       EK -> VR (IN REF. 2).
SQ   SEQUENCE   913 AA;  105343 MW;  CE5029F4EB51C1AA CRC64;

  Query Match             3.0%;  Score 153;  DB 1;  Length 913;
  Best Local Similarity   18.8%;  Pred. No. 0.21;
  Matches  175;  Conservative  125;  Mismatches  321;  Indels  312;  Gaps   43;

Qy        198 SDTNSEESRDNQQFLTHVKLAN-AKQTM---EDEQ------------------------ 228
              ||: ||  |    :: :|       |  ||  :    |||:
Db          9 SDSESESSLSGEELVTKPVGGNYGKQPLLLSEDEEDTKRVVRSAKDKRFEELTNLIRTIR 68

Qy        229 -GREARSHQKC-------GKACHPAEACA---GCQQEETDVVS--ESPLSDTGSEDVGTG 275
                :  |    ||       |||  |::       |   :     :::  |   |::   :  |
Db         69 NAMKIRDVTKCLEEFELLGKAYGKAKSIVDKEGVPRFYIRILADLEDYLNELWEDKEGKK 128

Qy        276 LKNANRLN-----RQESSLGNSPPFEK-----ESEPESPMDVDNSKNSCQDSEADEETSP 325
                | |         ||:     |    ||           :  ||    |  |   ||     | | | |
Db        129 KMNKNNAKALSTLRQKIRKYNR-DFESHITSYKQNPEQSADEDAEKN-----EEDSEGSS 182

Qy        326 GFDEQEDSSSAQTANKPSRFQPREADTELRKRSSAKGGEIR--LHFQFEGGESRAGMNDV 383
                || ||  || |                          |:|:|  |   ||  |    |       :   |
Db        183 DEDEDEDGVSAATF--------------LKKKSEAPSGESRKFLKKMDDEDEDSEDSEDD 228

Qy        384 NAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVP-------KAEDKRKEQCEMKHQRT 436
                 ||||     :  :  ::  |:: :    :       :    |        || ||::: :  || |
Db        229 EDWDTGSTSS---DSDSEEEEGKQTALASRFLKKAPTTDEDKKAAEKKREDKAKKKHDRK 285

Qy        437 ERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRP-------SANHTVTIRVDL 489
                             |::          |     |:  |  |:  |     :|              |  |  |:
Db        286 SKRL----------DEEEDNEGGEWERV-RGGVPLVKEKPKMFAKGTEITHAVVIK--- 331
              :::

Qy        490 LRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLT 549
              :: |:                            |        :  :|||  :  ||:|  || ::
Db        332 -KLNEI-------------------------LQARGKKGTDRAA--QIELLQ--LLVQIA 361

Qy        550 RPQNLKDAI---LKYNVAYSKKWDFTALI------DFWDKVLEEAEAQHLYQSILPDMVK 600
                 || :  :     :|:|:  |  :|:   :            :  | | |:          : :::
Db        362 AENNLGEGVIVKIKFNIIAS-LYDYNPNLATYMKPEMWGKCLD----------CINELMD 410

Qy        601 IALCLPNI---------------CTQPIPL------LKQKMNHSITM--------SQEQI 631
              |    |||                ||:  :          |  ::|:       |              ||| :
Db        411 ILFANPNIFVGENILEESENLHNADQPLRVRGCILTLVERMDEEFTKIMQNTDPHSQEYV 470

Qy        632 ASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLF-----------------------EG 668
               |  |   |     |   :   |   :  ::   |::                          ||
Db        471 EHLKDEAQVCAIIERVQRYLEEKGTTEEV--CRIYLLRILHTYYKFDYKAHQRQLTPPEG 528

Qy        669 RSSRKPEKLKT-------------LFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKL 715
               |  : :: :                  : |  :  |::  |   :                 |  :  |
Db        529 SSKSEQDQAENEGEDSAVLMERLCKYIYAKDRTDRIRTCAILCHIYHHALHSRWYQARDL 588

Qy        716 LTRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTS--AGLVQEEIRFLINPELIVSRLFT 773
               :   |:  :  |: :   :|: :    |   |: :   || ::
Db        589 MLMSHL--QDNIQ-HADPPVQILYNRTMVQLGICAFRQGLTKDA---------------- 629

Qy        774 EVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRY 833
                 || | |  : :  |  |        |  :            |: : :|:
Db        630 ----HNALLDIQSSGRAKELLGQGLLLRSLQERNQEQEKVERRRQ--------------- 670

Qy        834 LDQFVPEKIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQIL--AA 891
                  ||  :   |      | :|              :||:           |  |||  : : :
Db        671 ----VPFHLHINLELLECVYL--------VSAMLLEIPYMAAHESDARRRMISKQFHHQL 718

Qy        892 AVAERDVVYFTFGDSELMR-------------DIYSMHTFLTERKLTVGEVYKL------ 932
                | ||  :    |   | ||                     |  : |:|:   |:   |:|: |
Db        719 RVGERQPL---LGPPESMREHVVAASKAMKMGDWKTCHSFIINEKMN-GKVWDLFPEADK 774

Qy        933 ----LLRYYNEECRNCSTPGPDIKLYPFIYHAV 961
                  |:|   ||              :: | | | :|
Db        775 VRTMLVRKIQEE---------SLRTYLFTYSSV 798

RESULT 15
DPOZ_MOUSE
ID   DPOZ_MOUSE     STANDARD;      PRT;  3122 AA.
AC   Q61493; Q9JMD6; Q9QWX6;
DT   30-MAY-2000 (Rel. 39, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA polymerase zeta catalytic subunit (EC 2.7.7.7) (Seizure-related
DE   protein 4).
GN   REV3L OR POLZ OR SEZ4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Ola; TISSUE=Testis;
RX   MEDLINE=99202265; PubMed=10102037;
RA   Van Sloun P.P.H., Romeijn R.J., Eeken J.C.J.;
RT   "Molecular cloning, expression and chromosomal localisation of the
RT   mouse Rev3l gene, encoding the catalytic subunit of polymerase zeta.";
RL   Mutat. Res. 433:109-116(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kajiwara K.;
RT   "Molecular analyses of Sez4 encoding murine homologue of yeast REV3 in
RT   brain neurons.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 2368-3122 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic brain;
RX   MEDLINE=96216731; PubMed=8645260;
RA   Kajiwara K., Nagawawa H., Shimizu-Nishikawa K., Ookura T., Kimura M.,
RA   Sugaya E.;
RT   "Molecular characterization of seizure-related genes isolated by
RT   differential screening.";
RL   Biochem. Biophys. Res. Commun. 219:795-799(1996).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF083464; AAC98785.1; -.
DR   EMBL; AB031049; BAA90768.1; -.
DR   EMBL; D78644; BAA11461.1; -.
DR   PIR; T17202; T17202.
DR   MGD; MGI:1337131; Rev31.
DR   InterPro; IPR006172; DNA_pol_B.
```

```
DR   InterPro; IPR006134; DNA_pol_B_dom.
DR   InterPro; IPR006133; DNA_pol_B_exo.
DR   InterPro; IPR004578; Pol2.
DR   Pfam; PF00136; DNA_pol_B; 1.
DR   Pfam; PF03104; DNA_pol_B_exo; 1.
DR   PRINTS; PR00106; DNAPOLB.
DR   SMART; SM00486; POLBc; 1.
DR   TIGRFAMs; TIGR00592; pol2; 1.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   DNA-binding; DNA repair; Nuclear protein; Zinc-finger.
FT   ZN_FING      3034   3049       C4-TYPE (POTENTIAL).
FT   ZN_FING      3078   3096       C4-TYPE (POTENTIAL).
FT   CONFLICT       92     92       G -> A (IN REF. 2).
FT   CONFLICT      294    294       A -> T (IN REF. 2).
FT   CONFLICT      578    578       E -> Q (IN REF. 2).
FT   CONFLICT      609    609       R -> Q (IN REF. 2).
FT   CONFLICT     1278   1278       L -> P (IN REF. 2).
FT   CONFLICT     1298   1298       L -> F (IN REF. 2).
FT   CONFLICT     1416   1416       P -> L (IN REF. 2).
FT   CONFLICT     1848   1848       A -> T (IN REF. 2).
FT   CONFLICT     2368   2368       V -> G (IN REF. 3).
SQ   SEQUENCE   3122 AA;  350654 MW;  A39846CAF7365BA8 CRC64;

  Query Match             3.0%;  Score 153;  DB 1;  Length 3122;
  Best Local Similarity   19.4%;  Pred. No. 1;
  Matches  217;  Conservative  148;  Mismatches  387;  Indels  368;  Gaps   56;

Qy          2 SAGPGCEPCTKRPRW---DAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGC 58

Db       1122 SSPPRC--------WSPTDPRAEEIMAAAEKESMLFKGPNVYNTK--TVSPRVGKASRAR 1171

Qy         59 ALGRAGQHRGSATSLVF----KQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSV 114

Db       1172 AQVKKSKARLANSSVVTNKRNKRNQTTKLVDDGKKKPRAKQKQRANEKSLSRKHAIPAD- 1230

Qy        115 QKDNFYQHNMEKL----ENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSER-LL 169

Db       1231 --EKMKPHSEAELTPNHQSVSEL-----TSSSGAQALSKQK-----EMSQTGPAVDHPLL 1278

Qy        170 ESEPPAVTLVPEQFSNA---------NVD----QSSPKDDHSDT-----------NSEE 204

Db       1279 PAQPTGIS-AQQRLSNCFSSLLESKKSVDLRTFPSSRDDSHSSVVYSSIGPGISKINIQR 1337

Qy        205 SRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGKACHPAEACAGCQQEETDVVSESPL 264

Db       1338 SHNQSAMFTRKETTLIQKSIFDLSNHLSQVAQS-------TQVCSGIISPKTE-ESSSTQ 1389

Qy        265 SDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESEPESPMDVDNSKNS---CQDSEADE 321

Db       1390 KNCGS--------SMGKLNEYRSSLESKP--EQVCAPNFLHCKDSQQQTVSVSEQSKTSE 1439

Qy        322 ETSPGFDEQEDSSSAQTAN-----------------------------KPSRFQPR---E 349

Db       1440 TCSPGNAASEES---QTPNCFVTSLKSPIKQIAWEQKQRGFILDMSNFKPEKVKQRSLSE 1496

Qy        350 ADTELRKRSSAKGGEIRLHFQF-EGGESRAGMNDVNAKRPGSTSSLNV--ECRNSKQHGR 406

Db       1497 AISQTKALSQCKNQNVSTPSVFGEGQSGLAVLKELLQKRQQKAQSTNVVQDSTSTHQPDK 1556

Qy        407 KDSKITDHFMRVPKAEDKRKEQCEMKHQRTERKI-PKYIPPH------------------ 447

Db       1557 NISVSNEH----KKANKRTRPVTSPRKPRTPRRTKPKEQTPRRLKVDPLNLQTSGHLDNS 1612

Qy        448 LSPDKKWL-GTPIEEMRRMPRCGIRLPPLRPSANHTVTIRVDLLRIGE------------ 494

Db       1613 LSDDSPILFSDPGFE-----SCYSLEDSLSPEHNY----NFDINTIGQTGFCSFYSGSQF 1663

Qy        495 VP--KPFPTHF-----KDLWDNK---------HVKMPCSEQ--------------NLYP- 523

Db       1664 VPADQNLPQKFLSDAVQDLFPGQAIDKSELLSHDRQSCSEEKHHVSDSSPWIRASTLNPE 1723

Qy        524 ----VEDENGERAAGSRWE-------------------------LIQTALLNR------- 547

Db       1724 LFEKVAMDNNENHRHSQWKNSFHPLTSHSNSIMESFCVQQAENCLTEKSRLNRSSVSKEV 1783

Qy        548 -LTRPQ-NLKDAILKYN------VAYSKKWDFTALIDFWDKVLEEAEAQ--HLYQSILPD 597

Db       1784 FLSLPQANSSDWIQGHNRKEADQSLHSANTSFTTILSSPDGELVDAASEDLELYVSRNND 1843

Qy        598 MVKIAL-CLPNICTQPIPLLKQKMNHSIT----------MS----QEQIASLLANAF--- 639

Db       1844 VLTPAPDSSPRSTSSPL----QSKNGSFTPRTAHILKPLMSPPSREEIVATLLDHDLSEA 1899

Qy        640 -----FCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTG 694

Db       1900 IYQEPFCSNP-------------------------SDVPEKPREIGGRLLMVETRLPND 1933

Qy        695 LVTFTRQ-SLEDFPEWERCEKLLTRLHVTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGL 753

Db       1934 LIEFEGDFSLEGLRLWKTAFSAMTQ--NPRPGSPLRNGQAVVNKESSN--------SHKM 1983

Qy        754 VQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDRSERD 813

Db       1984 VEDK-KIVIMP----------------C-------KYAPSRQLVQAWLQAKEEYERSKK- 2018

Qy        814 DWQRRTTEIVAIDALHFRRYLDQFVPEKIRRELNKAYCGFLRPGVS-------------- 859

Db       2019 ---LPKTELTPVTK----------SAENVSPSLNPGDTCAVSPQVDKCPHTLSSSAHTKE 2065

Qy        860 --SENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERD 897

Db       2066 EVSKSQIALQTSTTGCS--------QTLLAAASAAVPEED 2097


Search completed: May 26, 2004, 18:41:04
Job time : 24.6533 secs
```

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 26, 2004, 15:28:37 ; Search time 173.272 Seconds
                                            (without alignments)
                                            13270.296 Million cell updates/sec

Title:           US-09-302-812-3
Perfect score:   7208
Sequence:        1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150052_5654/app_query.fasta_1.12437
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_414_@runat_26052004_150052_5654 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       A_Geneseq_29Jan04:*
                 1:  geneseqp1980s:*
                 2:  geneseqp1990s:*
                 3:  geneseqp2000s:*
                 4:  geneseqp2001s:*
                 5:  geneseqp2002s:*
                 6:  geneseqp2003as:*
                 7:  geneseqp2003bs:*
                 8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	72.0	976	5	AAE25630	Aae25630 Human pol
2	5190	72.0	976	5	AAU76021	Aau76021 Human pol
3	5190	72.0	976	5	AAU76012	Aau76012 Human pol
4	5190	72.0	976	6	ABG72279	Abg72279 Human pol
5	4635.5	64.3	977	5	AAE25629	Aae25629 Bovine po
6	4635.5	64.3	977	5	AAU76020	Aau76020 Bovine po
7	4635.5	64.3	977	5	AAU75799	Aau75799 Bovine po
8	4635.5	64.3	977	6	ABG72278	Abg72278 Bovine po
9	4395	61.0	968	5	AAE25631	Aae25631 Murine po
10	4395	61.0	968	5	AAU76022	Aau76022 Mouse pol
11	4395	61.0	968	5	AAU76013	Aau76013 Mouse pol
12	4395	61.0	968	6	ABG72280	Abg72280 Mmurine p
13	1081	15.0	768	4	ABB59491	Abb59491 Drosophil
14	1081	15.0	768	5	AAE25632	Aae25632 Fruit fly
15	1081	15.0	768	5	AAU76023	Aau76023 Fruit fly
16	1081	15.0	768	5	AAU76014	Aau76014 Fruit fly
17	1081	15.0	768	6	ABG72281	Abg72281 Fruit fly
18	463	6.4	726	5	AAE25633	Aae25633 Poly aden
19	463	6.4	726	5	AAU76024	Aau76024 Worm poly
20	463	6.4	726	5	AAU76015	Aau76015 Worm poly
21	463	6.4	726	6	ABG72282	Abg72282 C. elegan
22	344.5	4.8	819	4	ABG20721	Abg20721 Novel hum
23	210	2.9	100	4	ABG11103	Abg11103 Novel hum
24	178	2.5	33	5	AAE25652	Aae25652 Bovine po
25	177	2.5	1163	4	AAU28028	Aau28028 Novel hum
26	175	2.4	1111	4	AAU37362	Aau37362 Staphyloc
27	175	2.4	1274	6	ABJ18980	Abj18980 Pathogen
28	175	2.4	1274	6	ABM71209	Abm71209 Staphyloc
29	174	2.4	1274	6	ABU15913	Abu15913 Protein e
30	174	2.4	4873	6	ABO14747	Abo14747 Novel hum
31	173.5	2.4	748	4	AAM58340	Aam58340 Human bra
32	173	2.4	1948	5	ABP73774	Abp73774 Candida a
33	170.5	2.4	768	4	AAM78686	Aam78686 Human pro
34	170.5	2.4	2703	4	ABB60074	Abb60074 Drosophil
35	170	2.4	335	4	ABG09631	Abg09631 Novel hum
36	169.5	2.4	1803	4	ABB65391	Abb65391 Drosophil
37	169	2.3	1294	6	ABP75935	Abp75935 Human sec
38	168.5	2.3	2194	4	AAM40114	Aam40114 Human pol
39	168	2.3	1270	7	ADC37572	Adc37572 Human nuc
40	167	2.3	31	5	AAE25634	Aae25634 Bovine po
41	167	2.3	31	5	AAE25651	Aae25651 Bovine po
42	167	2.3	31	5	AAU76025	Aau76025 Bovine po
43	167	2.3	31	5	AAU76016	Aau76016 Bovine po
44	167	2.3	31	6	ABG72283	Abg72283 Oligopept
45	166	2.3	1163	6	ABU62602	Abu62602 Human maj

ALIGNMENTS

```
RESULT 1
AAE25630
ID   AAE25630 standard; protein; 976 AA.
XX
AC   AAE25630;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Human poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Homo sapiens.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42082.
```

```
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 55-60; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is human PARG
XX
SQ   Sequence 976 AA;

Alignment Scores:
Pred. No.:              0          Length:         976
Score:                  5190.00    Matches:        976
Percent Similarity:     100.00%    Conservative:   0
Best Local Similarity:  100.00%    Mismatches:     0
Query Match:            72.00%     Indels:         0
DB:                     5          Gaps:           0

US-09-302-812-3 (1-4069) x AAE25630 (1-976)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy    227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy    287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy    347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy    407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy    467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy    527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy    587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy    647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy    707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy    767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy    827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy    887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy    947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy   1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy   1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy   1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy   1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy   1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy   1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy   1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy   1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy   1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy   1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy   1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy   1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy   1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560
```

```
Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 2
AAU76021
ID   AAU76021 standard; protein; 976 AA.
XX
AC   AAU76021;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Homo sapiens.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14932.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 55-60; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the human PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 976 AA;

 Alignment Scores:
 Pred. No.:          0             Length:      976
 Score:              5190.00       Matches:     976
```

```
Percent Similarity:       100.00%      Conservative:  0
Best Local Similarity:    100.00%      Mismatches:    0
Query Match:              72.00%       Indels:        0
DB:                       5            Gaps:          0

US-09-302-812-3 (1-4069) x AAU76021 (1-976)

Qy     167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy     227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy     287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy     347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy     407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy     467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy     527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy     587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy     647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy     707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy     767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy     827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy     887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy     947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy    1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy    1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy    1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy    1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy    1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy    1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy    1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy    1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy    1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy    1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy    1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy    1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy    1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy    1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy    1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy    1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy    1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy    2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy    2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy    2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy    2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700
```

```
Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 3
AAU76012
ID   AAU76012 standard; protein; 976 AA.
XX
AC   AAU76012;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Homo sapiens.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14494.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 55-60; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the human PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 976 AA;

Alignment Scores:
Pred. No.:               0             Length:         976
Score:                   5190.00       Matches:        976
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             72.00%        Indels:         0
DB:                      5             Gaps:           0

US-09-302-812-3 (1-4069) x AAU76012 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120
```

```
Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 4
ABG72279
ID   ABG72279 standard; protein; 976 AA.
XX
AC   ABG72279;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14478.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents human PARG enzyme
XX
SQ   Sequence 976 AA;

Alignment Scores:
Pred. No.:              0            Length:         976
Score:                  5190.00      Matches:        976
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            72.00%       Indels:         0
DB:                     6            Gaps:           0

US-09-302-812-3 (1-4069) x ABG72279 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220
```

```
Qy    827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy    887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy    947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy   1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy   1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy   1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy   1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy   1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy   1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy   1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy   1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy   1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy   1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy   1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy   1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy   1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy   1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy   1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy   1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy   2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy   2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy   2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy   2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy   2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy   2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy   2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy   2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy   2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy   2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy   2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy   2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy   2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy   2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy   2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy   2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy   2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 5
AAE25629
ID   AAE25629 standard; protein; 977 AA.
XX
AC   AAE25629;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).
XX
KW   Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Bos taurus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42081.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 47-45; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is bPARG
XX
SQ   Sequence 977 AA;

Alignment Scores:
Pred. No.:               0          Length:          977
Score:                   4635.50    Matches:         874
Percent Similarity:      92.83%     Conservative:    32
Best Local Similarity:   89.55%     Mismatches:      69
Query Match:             64.31%     Indels:          1
DB:                      5          Gaps:            1

US-09-302-812-3 (1-4069) x AAE25629 (1-977)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              |||:::|||||||||||||||||||||||||||||||         ||||||   ||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              ||||||   |||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||   |||||||||   ||||||||||||||||||||||||||||      |||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              ||||||||||||||||||||||||||||||||||   |||||||||||||||   ||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
              |||||||||:::|||||||||   |||||||||||||||         |||||||||
Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||   ||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||   |||:::|||:::||||||||||||||||||         |||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              |||||||||||||||||||||:::|||:::|||         |||||||||||||||||
Db        181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              :::|||||||||:::||||||||||||||||||||||||||||   |||||||||||||||||
Db        201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
              |||||||||   |||||||||   |||||||||:::|||||||||||||||   |||
Db        221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              :::|||||||||   |||   ||||||   |||||||||:::|||   |||||||||
Db        241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
              :::|||||||||||||||   ||||||||||||||||||||||||   ||||||   |||
Db        261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
              :::|||   ||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
              ||||||||||||||||||||||||||||||||||||   ||||||:::|||||||||||||
Db        321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
              |||||||||||||||||   |||:::||||||   |||   |||||||||   |||
Db        341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360
```

```
Qy     1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db      361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy     1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db      381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy     1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db      401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy     1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db      421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy     1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db      441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy     1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db      461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy     1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db      481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy     1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db      501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy     1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db      521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy     1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db      541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy     1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db      561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy     1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db      581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy     1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db      601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy     2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db      621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy     2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db      641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy     2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db      661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy     2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db      681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy     2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db      701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy     2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db      721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy     2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db      741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy     2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db      761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy     2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db      781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy     2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db      801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy     2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db      821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy     2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db      841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy     2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db      861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy     2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db      881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy     2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db      901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy     2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db      921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy     2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db      941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy     3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
Db      961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

```
RESULT 6
AAU76020
ID   AAU76020 standard; protein; 977 AA.
XX
AC   AAU76020;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Region          601. .617
FT                   /note= "Represents PARG oligopeptide #1"
FT
```

```
FT   Region          761. .770
FT                   /note= "Represents PARG oligopeptide #2"
FT   Region          771. .801
FT                   /note= "Represents PARG oligopeptide #3"
FT   Region          849. .880
FT                   /note= "Represents PARG oligopeptide #4"
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14931.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 47-52; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the bovine PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 977 AA;

Alignment Scores:
Pred. No.:              0             Length:         977
Score:                  4635.50       Matches:        874
Percent Similarity:     92.83%        Conservative:   32
Best Local Similarity:  89.55%        Mismatches:     69
Query Match:            64.31%        Indels:         1
DB:                     5             Gaps:           1

US-09-302-812-3 (1-4069) x AAU76020 (1-977)

Qy      167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
            |||:::|||||||||||||||||||||||||||||||      ||||||   ||||||
Db        1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy      227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
            ||||||   ||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy      284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
            |||   |||||||||   |||||||||||||||||||||||||||||||      |||
Db       41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy      344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy      404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
            |||||||||||||||||||||||||||||||||||   ||||||||||||||   ||||||
Db       81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy      464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
            ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db      101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy      524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
            |||||||||:::|||||||||   |||||||||||||||         ||||||||
Db      121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy      584 GAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
            ||||||   |||||||||||   ||||||||||||||||||||||||||||||||||||
Db      141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy      644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
            ||||||   |||:::|||:::||||||||||||||||||      |||||||||||||||
Db      161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy      704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
            |||||||||||||||||||:::|||:::|||        |||||||||||||||||
Db      181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy      764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
            :::|||||||||:::||||||||||||||||||||||||   ||||||||||||||||||
Db      201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy      824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
            |||||||||   |||||||||      |||||||||:::|||||||||||||||   |||
Db      221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy      884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
            :::|||||||||   |||   ||||||   ||||||||||:::|||   |||||||||
Db      241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy      944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
            :::||||||||||||||   |||||||||||||||||||||||   ||||||   |||
Db      261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy     1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
            :::|||   |||||||||||||   ||||||||||||||||||||||||||||||||||||
Db      281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy     1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy     1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
            ||||||||||||||||||||||||||||||||||||   ||||||:::||||||||||||
Db      321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy     1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
            ||||||||||||||||||   |||:::||||||   |||   |||||||||   |||
Db      341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy     1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
            ||||||||||||:::|||||||||||||||||||||||||||||||||   ||||||
Db      361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy     1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
            ||||||:::|||||||||   ||||||:::   ||||||||||||||||||||||||||||
Db      381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy     1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
            |||||||||||||||:::||||||||||||||||||||||   ||||||:::||||||
Db      401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy     1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
            |||||||||:::||||||||||||   |||   |||||||||||||||||||||||||||
Db      421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy     1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
```

```
                 ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
              ||||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy       1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy       1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
              ||||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy       1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
              ||||||||||||||||:::   |||||||||||||||||||||||||||||||||||||||
Db        541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||   ||||||||||||||   :::||||||   |||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
              |||||||||||    :::|||    :::       ||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 7
AAU75799
ID   AAU75799 standard; protein; 977 AA.
XX
AC   AAU75799;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Bos taurus.
XX
FH   Key   Location/Qualifiers
FT   Region   601. .617
FT   /note= "Represents PARG oligopeptide #1"
FT   Region   761. .770
FT   /note= "Represents PARG oligopeptide #2"
FT   Region   771. .801
FT   /note= "Represents PARG oligopeptide #3"
FT   Region   849. .880
FT   /note= "Represents PARG oligopeptide #4"
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;

```
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14493.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 45-52; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the bovine PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 977 AA;

Alignment Scores:
Pred. No.:              0            Length:         977
Score:                  4635.50      Matches:        874
Percent Similarity:     92.83%       Conservative:   32
Best Local Similarity:  89.55%       Mismatches:     69
Query Match:            64.31%       Indels:         1
DB:                     5            Gaps:           1

US-09-302-812-3 (1-4069) x AAU75799 (1-977)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
Db      1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy    227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db     21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db     41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db    101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy    584 GAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
```

```
Db        541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||    ||||||||||||||    :::||||||   |||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              |||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
              |||||||||||    :::|||    :::       ||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

```
RESULT 8
ABG72278
ID   ABG72278 standard; protein; 977 AA.
XX
AC   ABG72278;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Bos taurus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14477.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
```

```
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents bovine PARG enzyme
XX
SQ   Sequence 977 AA;

Alignment Scores:
Pred. No.:              0           Length:         977
Score:                  4635.50     Matches:        874
Percent Similarity:     92.83%      Conservative:   32
Best Local Similarity:  89.55%      Mismatches:     69
Query Match:            64.31%      Indels:         1
DB:                     6           Gaps:           1

US-09-302-812-3 (1-4069) x ABG72278 (1-977)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
Db      1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy    227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db     21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db     41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db    101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
```

```
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::|||||||||||||   ||||||||||||||   :::||||||   |||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
              |||||||||||   :::|||   :::      |||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 9
AAE25631

```
ID   AAE25631 standard; protein; 968 AA.
XX
AC   AAE25631;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Mus musculus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42083.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 63-68; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is murine PARG
XX
SQ   Sequence 968 AA;

Alignment Scores:
Pred. No.:              0            Length:        968
Score:                  4395.00      Matches:       837
Percent Similarity:     90.99%       Conservative:  52
Best Local Similarity:  85.67%       Mismatches:    78
Query Match:            60.97%       Indels:        10
DB:                     5            Gaps:          6

US-09-302-812-3 (1-4069) x AAE25631 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||   ||||||||||||||||   ||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||::::::|||   |||||||||:::|||||||||||   ||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              ||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
```

```
Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db     97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db    712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy   2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db    732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy   2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db    752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy   2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
```

```
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              |||||:::|||||||||||||||||||||||||:::||||||    :::||||||||||||
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||   ||||||   |||||||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              |||||||||   |||||||||   |||      ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

RESULT 10
AAU76022
ID   AAU76022 standard; protein; 968 AA.
XX
AC   AAU76022;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Mus musculus.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998; 98US-0083768P.
PR   30-APR-1999; 99US-00302812.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14933.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 63-70; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the mouse PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 968 AA;

```
Alignment Scores:
Pred. No.:               0            Length:         968
Score:                   4395.00      Matches:        837
Percent Similarity:      90.99%       Conservative:   52
Best Local Similarity:   85.67%       Mismatches:     78
Query Match:             60.97%       Indels:         10
DB:                      5            Gaps:           6

US-09-302-812-3 (1-4069) x AAU76022 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||   ||||||||||||||||||   ||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||::::::|||   |||||||||:::||||||||||||   ||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTCAAACAAAAGACTATT 403
              |||:::||||||   |||||||||:::||||||||   ||||||||||||||||||||
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||:::||||||||||||||||||||   ||||||||||||||||||         ||||||
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                  |||   ||||||||||||   ||||||   ||||||:::||||||||||||   |||
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              |||||||||:::|||||||||||||||||||   |||||||||::::::||||||||||||
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||||||||   ||||||||||||   |||||||||         |||   :::|||
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              :::|||   ||||||||||||||||   :::||||||      ||||||||||||||||||
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196
```

```
Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216
Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233
Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253
Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273
Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293
Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313
Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332
Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352
Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371
Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391
Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411
Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431
Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451
Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471
Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491
Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511
Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531
Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551
Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571
Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591
Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611
Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631
Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651
Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691
Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711
Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db    712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
Qy   2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db    732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751
Qy   2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db    752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
Qy   2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db    772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791
Qy   2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db    792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811
Qy   2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db    812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831
Qy   2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db    832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851
Qy   2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db    852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871
Qy   2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db    872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891
Qy   2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db    892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
Qy   2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
```

```
                |||||||||||    ||||||   |||||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
               ||||||||   ||||||||   |||       ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 11
AAU76013
ID   AAU76013 standard; protein; 968 AA.
XX
AC   AAU76013;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Mus musculus.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14495.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 63-68; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the mouse PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 968 AA;

 Alignment Scores:
 Pred. No.:              0            Length:         968
 Score:                  4395.00      Matches:        837
 Percent Similarity:     90.99%       Conservative:   52
 Best Local Similarity:  85.67%       Mismatches:     78
 Query Match:            60.97%       Indels:         10
 DB:                     5            Gaps:           6

US-09-302-812-3 (1-4069) x AAU76013 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||   ||||||||||||||||   ||||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||:::::|||   |||||||||:::||||||||||||   ||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||   ||||||||||:::|||||||||   ||||||||||||||||||||
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||:::||||||||||||||||||   ||||||||||||||||||           ||||||
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                 |||   ||||||||||||   ||||||   ||||||:::|||||||||||   |||
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              |||||||||:::||||||||||||||||||||||||:::::::||||||||||||||||
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||||||||   ||||||||||||   |||||||||         |||   :::|||
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              :::|||   |||||||||||||||   :::||||||      |||||||||||||||||
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              |||   ||||||:::||||||||||||||||||||||||||   :::||||||||||||
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
              |||   |||         |||   ||||||         |||:::   |||||||||
Db        217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              |||         :::||||||      |||||||||:::|||:::||||||:::|||
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
              :::||||||||||||||||||||||:::||||||:::|||||||||   ||||||||||||
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
              ||||||      |||||||||   ||||||:::||||||||||||||||||||||||||
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
              |||||||||||||||   ||||||||||||:::|||         ||||||||||||||||
Db        314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
              |||   |||      |||||||||:::||||||   :::   ||||||||||||   |||
```

```
Db     333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352
Qy    1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db     353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371
Qy    1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db     372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391
Qy    1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db     392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411
Qy    1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db     412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431
Qy    1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db     432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451
Qy    1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db     452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471
Qy    1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db     472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491
Qy    1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db     492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511
Qy    1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db     512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531
Qy    1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db     532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551
Qy    1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db     552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571
Qy    1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db     572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591
Qy    1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db     592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611
Qy    2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db     612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631
Qy    2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db     632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651
Qy    2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db     652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
Qy    2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db     672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691
Qy    2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db     692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy    2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db     712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
Qy    2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db     732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751
Qy    2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db     752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
Qy    2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db     772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791
Qy    2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db     792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811
Qy    2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db     812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831
Qy    2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db     832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851
Qy    2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db     852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871
Qy    2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db     872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891
Qy    2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db     892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
Qy    2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db     912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931
Qy    2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db     932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951
Qy    3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
Db     952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

```
RESULT 12
ABG72280
ID   ABG72280 standard; protein; 968 AA.
XX
AC   ABG72280;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;
KW   cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;
KW   vasotropic; anticonvulsant; cerebroprotective; enzyme.
XX
OS   Mus musculus.
```

```
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14479.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents murine PARG enzyme
XX
SQ   Sequence 968 AA;

Alignment Scores:
Pred. No.:              0          Length:        968
Score:                  4395.00    Matches:       837
Percent Similarity:     90.99%     Conservative:  52
Best Local Similarity:  85.67%     Mismatches:    78
Query Match:            60.97%     Indels:        10
DB:                     6          Gaps:          6

US-09-302-812-3 (1-4069) x ABG72280 (1-968)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
          |||:::|||||||||||| |||||||||||||||| ||||||||||||||
Db      1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy    224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
          |||:::::::||| ||||||||:::||||||||||| ||||||||||||||||
Db     20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
          |||:::|||||| |||||||:::|||||||| ||||||||||||||||||||
Db     60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
          |||:::||||||||||||||||| |||||||||||||||||          ||||||
Db     80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
          |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db     97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
             |||  |||||||||||  ||||||  ||||||:::||||||||||||  |||
Db    117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
          |||||||||:::||||||||||||||||  |||||||||:::::||||||||||||||
Db    137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
          |||||||||||| |||||||||||| |||||||||       |||   :::|||
Db    157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
          :::|||  ||||||||||||||  :::||||||        ||||||||||||||||
Db    177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
          |||  ||||||:::||||||||||||||||||||||||||    :::||||||||||||
Db    197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
          |||  |||       |||       ||||||        |||:::  |||||||||
Db    217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
          |||             :::||||||      ||||||||:::|||:::|||||||:::|||
Db    234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
          :::||||||||||||||||||||:::||||||:::|||||||||   ||||||||||||
Db    254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
          ||||||        ||||||||||  ||||||:::||||||||||||||||||||||||||
Db    274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
          ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db    294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
          ||||||||||||||||   ||||||||||||:::|||      |||||||||||||||||
Db    314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
          |||  |||        ||||||||:::||||||    :::   ||||||||||||  |||
Db    333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
          ||||||   ||||||||||||||||||||||||||||      ||||||:::   |||
Db    353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
          :::|||||||||||||||     ||||||  |||||||||||||||||||||||||:::
Db    372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
          ||||||||||||||||:::||||||||||||||||||||||||  ||||||:::   |||
Db    392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
          :::||||||||||||||||||||||  |||   :::||||||||||||||||||||||||
Db    412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
          ||||||||:::||||||:::|||   |||:::|||||||||||||||||||||||||||
```

```
Db        432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
              ||||||||:::|||||||||||||||||   ||||||   ||||||||||||||||:::
Db        452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy       1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy       1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy       1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||||   ||||||:::||||||||||||||||||||||||
Db        712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||||||||:::||||||    :::|||||||||||
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              |||||||||   |||||||||   |||       ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

```
RESULT 13
ABB59491
ID   ABB59491 standard; protein; 768 AA.
XX
AC   ABB59491;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5265.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03594.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
```

```
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 768 AA;

Alignment Scores:
Pred. No.:              9.27e-89     Length:        768
Score:                  1081.00      Matches:       246
Percent Similarity:     57.89%       Conservative:  99
Best Local Similarity:  41.28%       Mismatches:    197
Query Match:            15.00%       Indels:        54
DB:                     4            Gaps:          16

US-09-302-812-3 (1-4069) x ABB59491 (1-768)

Qy     1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
Db       45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62
Qy     1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
Db       63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74
Qy     1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
Db       75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94
Qy     1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
Db       95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114
Qy     1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
Db      115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130
Qy     1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
Db      131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150
Qy     1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
Db      151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170
Qy     1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
Db      171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190
Qy     1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
Db      191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210
Qy     1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
Db      211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230
Qy     2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
Db      231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250
Qy     2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
Db      251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270
Qy     2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
Db      271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290
Qy     2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
Db      291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310
Qy     2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
Db      311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330
Qy     2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
Db      331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350
Qy     2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db      351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370
Qy     2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
Db      371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390
Qy     2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
Db      391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410
Qy     2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
Db      411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430
Qy     2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
Db      431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450
Qy     2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db      451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470
Qy     2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
Db      471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490
Qy     2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
Db      491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502
Qy     2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
Db      503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522
Qy     2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
Db      523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542
Qy     2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
Db      543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561
Qy     2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
Db      562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
Qy     3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
Db      582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598
Qy     3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
Db      599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614

RESULT 14
AAE25632
ID   AAE25632 standard; protein; 768 AA.
XX
AC   AAE25632;
XX
```

```
DT   04-NOV-2002  (first entry)
XX
DE   Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy; fruit fly.
XX
OS   Drosophila melanogaster.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42084.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Example 13; Col 71-76; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is fruit fly PARG
XX
SQ   Sequence 768 AA;

Alignment Scores:
Pred. No.:              9.27e-89      Length:        768
Score:                  1081.00       Matches:       246
Percent Similarity:     57.89%        Conservative:  99
Best Local Similarity:  41.28%        Mismatches:    197
Query Match:            15.00%        Indels:        54
DB:                     5             Gaps:          16

US-09-302-812-3 (1-4069) x AAE25632 (1-768)

Qy       1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
              |||:::   |||:::   |||             :::   ||||||:::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
                                                |||   :::|||      |||   |||      |||
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
                 |||      :::|||   :::         |||   |||               |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
              |||   :::               |||   |||      ::::::   :::
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy       1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
                    |||:::|||:::   :::         |||      ||||||::::::||||||
Db        115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy       1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
              ::::::||||||:::   ::::::   ||||||   |||:::   :::|||      |||
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
                    |||||||||:::|||:::   |||||||||                     :::
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
                 |||:::   ||||||:::   ||||||      |||      :::|||   |||
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
              ||||||            |||:::   |||:::|||:::|||   :::      ::::::
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
                 :::||||||   ::::::::::::||||||   ||||||::::::         |||
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
              :::||||||||||||         |||   |||:::::::::||||||:::||||||:::
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
                 ||||||||||||||||||   |||||||||||||||||||||      |||   |||
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
              ||||||||||||::::::||||||||||||||||||||||||||||::::::
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
                    ||||||||||||   :::      ||||||||||||||||      ||||||
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
              :::                     |||||||||:::|||||||||      |||:::|||
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
              |||         :::|||   :::            ||||||         |||||||||
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                    ||||||||||||||:::   |||   |||:::||||||||||||||||||||
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              :::::::::||||||||||||               |||   ||||||||||||||||:::
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              |||   |||||||||::::::   :::||||||||||||   |||         |||
Db        391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              |||::::::   |||   |||:::||||||   |||||||||||||||   ::::::
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||   :::   ||||||   :::   ||||||            ||||||   |||
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450
```

```
Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||     :::          |||:::      :::
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
              |||   ||||||||||||||||||||   ||||||:::
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
              ||||||                        ||||||||||||||||||||||||||||||
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2789 TTTGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
              ||||||||||||:::   ||||||||||||:::|||::::::        |||
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
              |||   :::   |||:::|||||||||:::   |||          |||   :::   |||
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy       2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
                   :::                   ||||||      ::::::      :::||||||
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
                 |||:::         :::      |||:::      |||
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy       3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
                    |||||||||   ||||||               |||          :::   |||
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
                    |||:::      :::         |||:::   |||:::|||
Db        599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614
```

```
RESULT 15
AAU76023
ID   AAU76023 standard; protein; 768 AA.
XX
AC   AAU76023;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Fruit fly poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Fruit fly; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Drosophila melanogaster.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14934.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Example 13; Col 71-76; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the fruit fly PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 768 AA;

Alignment Scores:
Pred. No.:              9.27e-89       Length:         768
Score:                  1081.00        Matches:        246
Percent Similarity:     57.89%         Conservative:   99
Best Local Similarity:  41.28%         Mismatches:     197
Query Match:            15.00%         Indels:         54
DB:                     5              Gaps:           16

US-09-302-812-3 (1-4069) x AAU76023 (1-768)
```

```
Qy       1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
                 |||:::   |||:::   |||                  :::   ||||||:::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
                                    |||   :::|||         |||   |||      |||
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
                 |||      :::|||   :::            |||   |||            |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
                 |||   :::               |||   |||      ::::::   :::
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy       1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
                    |||:::|||:::   :::         |||         ||||||::::::||||||
Db        115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy       1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
                 ::::::||||||:::   ::::::   ||||||   |||:::   :::|||      |||
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
                    ||||||||||:::|||:::   |||||||||                  :::
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
                 |||:::   ||||||:::   ||||||      |||         :::|||   |||
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
                 ||||||                  |||:::   |||:::|||:::|||   :::   ::::::
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
                 :::||||||   ::::::::::::||||||   ||||||::::::         |||
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
```

```
                  :::|||||||||||        |||     |||:::::::::||||||:::||||||:::
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
                 |||||||||||||||||    ||||||||||||||||||||||          |||    |||
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
              |||||||||||::::::||||||||||||||||||||||||||::::::
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
                           ||||||||||||    :::      |||||||||||||||      ||||||
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
              :::                   |||||||||:::|||||||||    |||:::|||
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2273 GAAGATTTTCCAGAATGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
              |||          :::|||    :::              ||||||          |||||||||
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                    ||||||||||||||:::    |||    |||:::||||||||||||||||||||
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              :::::::::||||||||||||          |||    ||||||||||||||||||||:::
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              |||    ||||||||::::::    :::|||||||||||    |||          |||
Db        391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              |||::::::    |||    |||:::||||||    ||||||||||||||||     ::::::
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||    :::    ||||||    :::    ||||||          ||||||    |||
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              |||||||||||||||||||||||||||     :::          |||:::     :::
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
              |||    |||||||||||||||||||||     ||||||:::
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
              ||||||                        |||||||||||||||||||||||||||||||
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
              ||||||||||||:::    ||||||||||||:::|||::::::       |||
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
              |||    :::    |||:::|||||||||:::    |||          |||    :::    |||
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy       2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
                 :::                   ||||||    ::::::    :::||||||
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
                 |||:::          :::    |||:::    |||
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy       3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
                    |||||||||    ||||||                |||          :::       |||

Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
                       |||:::       :::          |||:::    |||:::|||
Db        599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614
```

Search completed: May 26, 2004, 16:47:30
Job time : 307.605 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 26, 2004, 16:27:48 ; Search time 34.0417 Seconds
                                          (without alignments)
                                          12341.705 Million cell updates/sec

Title:           US-09-302-812-3
Perfect score:   7208
Sequence:        1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150054_5717/app_query.fasta_1.12437
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_131_@runat_26052004_150054_5717 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	72.0	976	4	US-09-302-812-4	Sequence 4, Appli
2	5190	72.0	976	4	US-09-511-477-4	Sequence 4, Appli
3	5190	72.0	976	4	US-09-511-507-4	Sequence 4, Appli
4	4635.5	64.3	977	4	US-09-302-812-2	Sequence 2, Appli
5	4635.5	64.3	977	4	US-09-511-477-2	Sequence 2, Appli
6	4635.5	64.3	977	4	US-09-511-507-2	Sequence 2, Appli
7	4395	61.0	968	4	US-09-302-812-6	Sequence 6, Appli
8	4395	61.0	968	4	US-09-511-477-6	Sequence 6, Appli
9	4395	61.0	968	4	US-09-511-507-6	Sequence 6, Appli
10	1081	15.0	768	4	US-09-302-812-8	Sequence 8, Appli
11	1081	15.0	768	4	US-09-511-477-8	Sequence 8, Appli
12	1081	15.0	768	4	US-09-511-507-8	Sequence 8, Appli
13	463	6.4	726	4	US-09-302-812-10	Sequence 10, Appl
14	463	6.4	726	4	US-09-511-477-10	Sequence 10, Appl
15	463	6.4	726	4	US-09-511-507-10	Sequence 10, Appl
16	167	2.3	31	4	US-09-302-812-11	Sequence 11, Appl
17	167	2.3	31	4	US-09-511-477-11	Sequence 11, Appl
18	167	2.3	31	4	US-09-511-507-11	Sequence 11, Appl
19	154	2.1	1972	4	US-08-875-435B-3	Sequence 3, Appli
20	152	2.1	1231	4	US-09-595-684B-23	Sequence 23, Appl
21	150.5	2.1	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
22	150.5	2.1	1972	4	US-08-875-435B-4	Sequence 4, Appli
23	150	2.1	3969	3	US-08-061-376-5	Sequence 5, Appli
24	149.5	2.1	1187	1	US-08-320-559-28	Sequence 28, Appl
25	149.5	2.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
26	149.5	2.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl
27	149.5	2.1	1210	1	US-08-320-559-26	Sequence 26, Appl
28	149.5	2.1	1210	3	US-08-545-860D-26	Sequence 26, Appl
29	149.5	2.1	1210	5	PCT-US94-04496-26	Sequence 26, Appl
30	149	2.1	26	4	US-09-302-812-20	Sequence 20, Appl
31	149	2.1	26	4	US-09-511-477-20	Sequence 20, Appl
32	149	2.1	26	4	US-09-511-507-20	Sequence 20, Appl
33	149	2.1	29	4	US-09-302-812-12	Sequence 12, Appl
34	149	2.1	29	4	US-09-511-477-12	Sequence 12, Appl
35	149	2.1	29	4	US-09-511-507-12	Sequence 12, Appl
36	149	2.1	427	4	US-09-134-001C-5143	Sequence 5143, Ap
37	148.5	2.1	1596	4	US-08-978-277A-4	Sequence 4, Appli
38	148	2.1	3248	1	US-08-353-700-1	Sequence 1, Appli
39	148	2.1	3248	5	PCT-US95-16216-1	Sequence 1, Appli
40	147	2.0	754	4	US-09-976-594-375	Sequence 375, App
41	146.5	2.0	2125	4	US-09-919-172-29	Sequence 29, Appl
42	146	2.0	1232	4	US-09-592-054-2	Sequence 2, Appli
43	145.5	2.0	913	4	US-08-971-089-4	Sequence 4, Appli
44	145.5	2.0	2482	1	US-08-328-254-6	Sequence 6, Appli
45	144.5	2.0	1085	1	US-08-431-080-28	Sequence 28, Appl

ALIGNMENTS

```
RESULT 1
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-302-812-4

Alignment Scores:
Pred. No.:               0              Length:        976
Score:                   5190.00        Matches:       976
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             72.00%         Indels:        0
DB:                      4              Gaps:          0

US-09-302-812-3 (1-4069) x US-09-302-812-4 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740
```

```
Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 2
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-477-4

Alignment Scores:
Pred. No.:                 0                Length:          976
Score:                     5190.00          Matches:         976
Percent Similarity:        100.00%          Conservative:    0
Best Local Similarity:     100.00%          Mismatches:      0
Query Match:               72.00%           Indels:          0
DB:                        4                Gaps:            0

US-09-302-812-3 (1-4069) x US-09-511-477-4 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 3
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
```

```
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-507-4

Alignment Scores:
Pred. No.:              0            Length:         976
Score:                  5190.00      Matches:        976
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            72.00%       Indels:         0
DB:                     4            Gaps:           0

US-09-302-812-3 (1-4069) x US-09-511-507-4 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620
```

```
Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

RESULT 4
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-302-812-2

Alignment Scores:			
Pred. No.:	0	Length:	977
Score:	4635.50	Matches:	874
Percent Similarity:	92.83%	Conservative:	32
Best Local Similarity:	89.55%	Mismatches:	69
Query Match:	64.31%	Indels:	1
DB:	4	Gaps:	1

US-09-302-812-3 (1-4069) x US-09-302-812-2 (1-977)

```
Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              |||:::|||||||||||||||||||||||||||           ||||||   ||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              ||||||   ||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||   ||||||||||   |||||||||||||||||||||||||||||    |||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||||||||||||||||||||||||||||||||   ||||||||||||||   ||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
              |||||||||:::|||||||||   |||||||||||||||||         |||||||||
Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||   ||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||   |||:::|||:::||||||||||||||||||         |||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              ||||||||||||||||||||||:::||||:::||||           ||||||||||||||||||
```

```
Db    181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db    721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy   2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db    741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy   2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db    761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy   2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db    781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy   2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db    801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy   2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db    821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy   2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db    841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy   2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db    861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy   2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db    881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy   2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db    901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920
```

```
Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
              ||||||||||||   :::|||    :::        ||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 5
US-09-511-477-2
; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-511-477-2

Alignment Scores:
Pred. No.:              0              Length:         977
Score:                  4635.50        Matches:        874
Percent Similarity:     92.83%         Conservative:   32
Best Local Similarity:  89.55%         Mismatches:     69
Query Match:            64.31%         Indels:         1
DB:                     4              Gaps:           1

US-09-302-812-3 (1-4069) x US-09-511-477-2 (1-977)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              |||:::||||||||||||||||||||||||||||||         ||||||   ||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              ||||||   |||||||||||||||||||||||||||||     ||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||   |||||||||   |||||||||||||||||||||||||||         |||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||   ||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
              |||||||||:::||||||||   |||||||||||||||         |||||||||

Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||   ||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||   |||:::|||:::||||||||||||||||||         ||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              |||||||||||||||||||:::|||:::|||         |||||||||||||||||
Db        181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              :::|||||||||:::|||||||||||||||||||||||||   |||||||||||||||||
Db        201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
              |||||||||   |||||||||         |||||||||:::||||||||||||||   |||
Db        221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              :::||||||||||   |||   ||||||   |||||||||:::|||   |||||||||
Db        241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
              :::|||||||||||||||   ||||||||||||||||||||||||   ||||||   |||
Db        261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
              :::|||   ||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
              |||||||||||||||||||||||||||||||||||||||   ||||||:::|||||||||||
Db        321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
              ||||||||||||||||||   |||:::||||||   |||   |||||||||   |||
Db        341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy       1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
              ||||||||||:::|||||||||||||||||||||||||||||||||||   ||||||
Db        361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy       1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
              |||||:::|||||||||   ||||||:::   |||||||||||||||||||||||||||
Db        381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy       1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
              ||||||||||||||||:::|||||||||||||||||||||||||   ||||||:::||||||
Db        401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy       1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
              |||||||||:::||||||||||||   |||   ||||||||||||||||||||||||||
Db        421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy       1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500
```

```
Qy     1664 ACACATTATAAAGATTTGTGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db      501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy     1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db      521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy     1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db      541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy     1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db      561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy     1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db      581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy     1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db      601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy     2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db      621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy     2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db      641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy     2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db      661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy     2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db      681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy     2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db      701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy     2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db      721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy     2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db      741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy     2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db      761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy     2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db      781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy     2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db      801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy     2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db      821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy     2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db      841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy     2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db      861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy     2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db      881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy     2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db      901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy     2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db      921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy     2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db      941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy     3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
Db      961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 6
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-2

Alignment Scores:			
Pred. No.:	0	Length:	977
Score:	4635.50	Matches:	874
Percent Similarity:	92.83%	Conservative:	32
Best Local Similarity:	89.55%	Mismatches:	69
Query Match:	64.31%	Indels:	1
DB:	4	Gaps:	1

US-09-302-812-3 (1-4069) x US-09-511-507-2 (1-977)

```
Qy      167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
Db        1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy      227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db       21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy      284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db       41 AspSerLysAspAlaProValGlnPheArgValProProSerSerGlyCysAlaLeu 60

Qy      344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
```

```
Db     61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db    101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db    721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy   2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db    741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy   2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db    761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy   2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db    781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800
```

```
Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 7
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

Alignment Scores:

Pred. No.:	0	Length:	968
Score:	4395.00	Matches:	837
Percent Similarity:	90.99%	Conservative:	52
Best Local Similarity:	85.67%	Mismatches:	78
Query Match:	60.97%	Indels:	10
DB:	4	Gaps:	6

US-09-302-812-3 (1-4069) x US-09-302-812-6 (1-968)

```
Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db        217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db        314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db        333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy       1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db        353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371
```

```
Qy      1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db       372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy      1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db       392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy      1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db       412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy      1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db       432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy      1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db       452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy      1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db       472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy      1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db       492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy      1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db       512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy      1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db       532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy      1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db       552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy      1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db       572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy      1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db       592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy      2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db       612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy      2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db       632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy      2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db       652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy      2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db       672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy      2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db       692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy      2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db       712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy      2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db       732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy      2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db       752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy      2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db       772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy      2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db       792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy      2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db       812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy      2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db       832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy      2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db       852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy      2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db       872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy      2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db       892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy      2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db       912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy      2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db       932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy      3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
Db       952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

RESULT 8
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-511-477-6

```
Alignment Scores:
Pred. No.:                0              Length:         968
Score:                    4395.00        Matches:        837
Percent Similarity:       90.99%         Conservative:   52
Best Local Similarity:    85.67%         Mismatches:     78
Query Match:              60.97%         Indels:         10
DB:                       4              Gaps:           6

US-09-302-812-3 (1-4069) x US-09-511-477-6 (1-968)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
Db      1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy    224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db     20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db     40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db     97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
```

```
Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||||   ||||||:::||||||||||||||||||||||||
Db        712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||:::||||||   :::||||||||||||||||||
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              |||||||||   |||||||||   |||      ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

RESULT 9
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

```
Alignment Scores:
Pred. No.:                0              Length:          968
Score:                    4395.00        Matches:         837
Percent Similarity:       90.99%         Conservative:    52
Best Local Similarity:    85.67%         Mismatches:      78
Query Match:              60.97%         Indels:          10
DB:                       4              Gaps:            6

US-09-302-812-3 (1-4069) x US-09-511-507-6 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||   ||||||||||||||||||   |||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||::::::|||   |||||||||:::||||||||||||   ||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||   |||||||||:::|||||||||   |||||||||||||||||||||
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||:::||||||||||||||||||   ||||||||||||||||||         ||||||
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                 |||   ||||||||||||   ||||||   ||||||:::||||||||||||   |||
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||:::|||||||||||||||   |||||||||::::::||||||||||||||||||
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||||||||   ||||||||||||   |||||||||         |||   :::|||
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              :::|||   |||||||||||||||   :::||||||   |||||||||||||||||||||
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              |||   ||||||:::||||||||||||||||||||||||   :::|||||||||||||||
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
              |||   |||      |||      ||||||         |||:::   |||||||||
Db        217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              |||            :::||||||   |||||||||:::|||:::||||||:::||||||
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253
```

```
Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
              :::||||||||||||||||||:::||||||:::|||||||||   ||||||||||||
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
              ||||||      |||||||||   ||||||:::|||||||||||||||||||||||||
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
              ||||||||||||||   ||||||||||||:::|||      |||||||||||||||||
Db        314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
              |||   |||      ||||||||:::||||||    :::    ||||||||||||  |||
Db        333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy       1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
              ||||||    |||||||||||||||||||||||||||      ||||||:::      |||
Db        353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy       1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
              :::|||||||||||||||        ||||||    |||||||||||||||||||||:::
Db        372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy       1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
              ||||||||||||||||||:::||||||||||||||||||||   ||||||:::   |||
Db        392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy       1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
              :::||||||||||||||||||||   |||    :::||||||||||||||||||||||
Db        412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy       1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
              |||||||||:::||||||:::|||   |||:::||||||||||||||||||||||||||
Db        432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
              |||||||||:::|||||||||||||||||   ||||||   ||||||||||||||||:::
Db        452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy       1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy       1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy       1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||      ||||||:::||||||||||||||||||||||||
Db        712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||||||||:::||||||    :::||||||||||||
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::|||||||||||||||||||||||||||||||||||||||||||||   |||
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||      ||||||      ||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              |||||||||   |||||||||   |||         ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968
```

RESULT 10

```
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-302-812-8

Alignment Scores:
Pred. No.:              4.06e-102     Length:        768
Score:                  1081.00       Matches:       246
Percent Similarity:     57.89%        Conservative:  99
Best Local Similarity:  41.28%        Mismatches:    197
Query Match:            15.00%        Indels:        54
DB:                     4             Gaps:          16

US-09-302-812-3 (1-4069) x US-09-302-812-8 (1-768)

Qy   1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
Db     45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy   1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
Db     63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy   1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
Db     75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy   1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
Db     95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy   1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
Db    115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy   1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
Db    131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy   1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
Db    151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy   1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
Db    171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy   1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
Db    191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy   1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
Db    211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy   2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
Db    231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy   2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
Db    251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy   2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
Db    271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy   2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
Db    291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy   2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
Db    311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy   2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
Db    331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy   2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db    351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy   2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
Db    371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy   2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
Db    391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy   2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
Db    411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy   2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
Db    431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy   2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db    451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy   2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
Db    471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy   2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
Db    491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy   2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
Db    503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy   2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
Db    523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy   2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
Db    543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy   2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
Db    562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
```

```
Qy       3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
                   |||||||||   ||||||           |||         :::     |||
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
                      |||:::      :::       |||:::   |||:::|||
Db        599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614

RESULT 11
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-511-477-8

Alignment Scores:
Pred. No.:              4.06e-102      Length:          768
Score:                  1081.00        Matches:         246
Percent Similarity:     57.89%         Conservative:    99
Best Local Similarity:  41.28%         Mismatches:      197
Query Match:            15.00%         Indels:          54
DB:                     4              Gaps:            16

US-09-302-812-3 (1-4069) x US-09-511-477-8 (1-768)

Qy       1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
               |||:::   |||:::   |||               :::   ||||||:::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
                                      |||   :::|||      |||   |||      |||
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
                 |||      :::|||   :::            |||   |||            |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
               |||   :::                  |||   |||      ::::::   :::
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy       1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
                       |||:::|||:::   :::      |||         ||||||::::::||||||
Db        115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy       1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
               ::::::||||||:::   ::::::   ||||||   |||:::   :::|||      |||
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
                          |||||||||:::|||:::   |||||||||                  :::
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
                 |||:::   ||||||:::   ||||||         |||         :::|||   |||
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
              ||||||               |||:::   |||:::|||:::|||   :::      ::::::
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
                 :::||||||   ::::::::::::||||||   ||||||::::::         |||
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
              :::||||||||||||         |||   |||:::::::::||||||:::||||||:::
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
                 ||||||||||||||||||   ||||||||||||||||||||||      |||   |||
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
              ||||||||||||::::::||||||||||||||||||||||||||||::::::
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
                          ||||||||||||   :::   |||||||||||||||      ||||||
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
              :::                  ||||||||:::|||||||||   |||:::|||
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2273 GAAGATTTTCCAGAATGGAAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
              |||          :::|||   :::         ||||||         |||||||||
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                    |||||||||||||||:::   |||   |||:::|||||||||||||||||||||
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ::::::::||||||||||||         |||   |||||||||||||||||||||:::
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              |||   ||||||||::::::   :::||||||||||||   |||         |||
Db        391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              |||::::::   |||   |||:::||||||   |||||||||||||||   ::::::
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||   :::   ||||||   :::   ||||||         ||||||   |||
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||         :::         |||:::   :::
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
              |||   |||||||||||||||||||||      ||||||:::
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
              ||||||                        ||||||||||||||||||||||||||||
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
              ||||||||||||:::   ||||||||||||:::|||::::::         |||
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
              |||   :::   |||:::|||||||||:::   |||         |||   :::   |||
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542
```

```
Qy     2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
Db      543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy     2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA----------------- 3010
Db      562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy     3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
Db      582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy     3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
Db      599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614
```

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Alignment Scores:

Pred. No.:	4.06e-102	Length:	768
Score:	1081.00	Matches:	246
Percent Similarity:	57.89%	Conservative:	99
Best Local Similarity:	41.28%	Mismatches:	197
Query Match:	15.00%	Indels:	54
DB:	4	Gaps:	16

US-09-302-812-3 (1-4069) x US-09-511-507-8 (1-768)

```
Qy     1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
Db       45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy     1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
Db       63 -----------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy     1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
Db       75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy     1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
Db       95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy     1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
Db      115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy     1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
Db      131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy     1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
Db      151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy     1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
Db      171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy     1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
Db      191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy     1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
Db      211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy     2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
Db      231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250

Qy     2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
Db      251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy     2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
Db      271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy     2177 AGGAAACCGGAGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
Db      291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy     2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
Db      311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy     2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
Db      331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy     2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db      351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy     2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
Db      371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy     2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
Db      391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy     2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
Db      411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy     2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
Db      431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy     2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db      451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy     2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
Db      471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy     2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
```

```
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
              ||||||||||||:::   ||||||||||||:::|||::::::         |||
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
              |||   :::   |||:::|||||||||:::   |||          |||   :::   |||
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy       2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
                 :::                     ||||||    ::::::       :::||||||
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
                 |||:::        :::      |||:::     |||
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy       3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
                    |||||||||   ||||||            |||          :::      |||
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
                    |||:::      :::         |||:::   |||:::|||
Db        599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614
```

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-10

Alignment Scores:

Pred. No.:	2.73e-38	Length:	726
Score:	463.00	Matches:	207
Percent Similarity:	38.11%	Conservative:	123
Best Local Similarity:	23.90%	Mismatches:	278
Query Match:	6.42%	Indels:	258
DB:	4	Gaps:	37

US-09-302-812-3 (1-4069) x US-09-302-812-10 (1-726)

```
Qy        599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
              :::   :::|||         |||         |||           |||      |||
Db         27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              |||      :::                        ||||||:::         :::
Db         47 GluSerLysGluAsp---------------------------ProGluGluProLysSer 57

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
                 ::::::      |||   |||:::   |||   |||      ||||||   ||||||
Db         58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy        779 --------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                 :::|||   :::     ||||||    |||:::|||:::
Db         78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy        815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
              :::|||   :::               |||         ||||||         :::
Db         98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy        863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
              :::|||               ||||||      |||   |||:::
Db        116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy        923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
                       ||||||      ||||||   :::                  ||||||
Db        126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy        983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
                 |||      |||::::::|||:::   :::|||:::
Db        137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy       1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
                    ||||||                   |||   :::         :::
Db        151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy       1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
                 ||||||      ||||||::::::            ||||||      :::
Db        164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy       1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                                   |||      :::||||||   |||
Db        184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy       1196 TTCCAAGCAAGAGACGCTGACATTGAATTT---------------------AGGAAACGG 1234
                 :::         :::||||||   |||                        |||
Db        204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy       1235 TACTCTACTAAGGGC--------------------------------------------- 1249
                    |||
Db        224 ValArgThrSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy       1250 GGTGAAGTTAGATTACAT------------------------TTCCAATTTGAAGGAGGA 1285
              |||      |||::::::                                 |||   |||
Db        244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy       1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
                 :::   |||         |||||||||        |||
Db        264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy       1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
                          |||      :::      :::   |||            ::::::|||
Db        283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy       1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
                 |||                        :::   ||||||            :::|||
Db        303 ProLys------------------------IleValLeuPro------------GlnArg 310

Qy       1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
                    :::::::::::::::   :::|||   :::                  ||||||
Db        311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy       1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
              :::      |||:::         :::   :::                        :::
Db        331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy       1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
                    |||               |||         :::||||||   |||
Db        351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy       1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657

Db        368 ------------------------------------------------------------ 368

Qy       1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
```

```
                 :::||||||:::::
Db        369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy       1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
                                                                       |||
Db        375 -------------------------------------------------------SerGlu 376

Qy       1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
              |||:::   :::     :::|||  ||||||  |||           :::
Db        377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy       1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTC 1894

Db        394 ------------------------------------------------------------ 394

Qy       1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                    |||    |||:::         :::  :::|||:::    ::::::   |||
Db        395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy       1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
              :::
Db        412 Val--------------------------------------------------------- 412

Qy       2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
                          |||  |||:::|||  :::|||  |||:::|||:::|||
Db        413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy       2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
                 ||||||                      |||:::    |||   :::|||
Db        429 MetPhePhe---------------------AlaArgProAspSerProPheSer------ 439

Qy       2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                          |||  |||:::           |||        |||||||||
Db        440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy       2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
              |||   ||||||   ||||||   :::::::::         |||   |||   |||:::
Db        456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy       2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
              |||      |||:::         :::   |||      ||||||:::      :::|||
Db        476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy       2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
                 |||      |||              ::::::   :::   |||:::
Db        494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy       2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
                    ||||||||||||||||||      :::||||||||||||   :::   |||
Db        514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy       2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
                 |||||||||||||||||||||:::   ||||||::::::::        |||
Db        534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy       2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
                 |||   :::         |||   :::   |||   |||         :::|||
Db        554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573

Qy       2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
              ||||||||||||      |||   :::|||                :::   |||   :::
Db        574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy       2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
              :::   :::|||   ||||||            |||      |||   :::|||||||||
Db        594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy       2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
              |||:::   |||:::                     :::|||
Db        614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              :::   ||||||:::   ||||||         ||||||:::      |||
Db        634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              |||:::      :::   ||||||   ||||||||||||||||||   ||||||
Db        654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||   :::||||||::::::||||||          |||:::|||   :::   :::
Db        673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::||||||:::   ||||||                 :::         :::   :::
Db        693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy       2927 AGGAAACTCACTGTTGGA 2944
              :::   :::|||:::|||
Db        713 LysAspValThrLeuGly 718
```

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI\
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-477-10

Alignment Scores:			
Pred. No.:	2.73e-38	Length:	726
Score:	463.00	Matches:	207
Percent Similarity:	38.11%	Conservative:	123
Best Local Similarity:	23.90%	Mismatches:	278
Query Match:	6.42%	Indels:	258
DB:	4	Gaps:	37

US-09-302-812-3 (1-4069) x US-09-511-477-10 (1-726)

```
Qy        599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
              :::   :::|||         |||               |||         |||   |||
Db         27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              |||         :::                         ||||||:::       :::
Db         47 GluSerLysGluAsp--------------------------ProGluGluProLysSer 57

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
                 ::::::      |||   |||:::   |||   |||      ||||||   ||||||
Db         58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy        779 ---------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                   :::|||   :::      ||||||   |||:::|||:::
Db         78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy        815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
              :::|||   :::                  |||         ||||||      :::
Db         98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115
```

```
Qy    863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
          :::|||            ||||||      |||   |||:::
Db    116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy    923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
                   ||||||      ||||||   :::                  ||||||
Db    126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy    983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
                |||      |||::::::|||:::   :::|||:::
Db    137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy   1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
                   ||||||                          |||   :::      :::
Db    151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy   1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
             ||||||      ||||||::::::               ||||||      :::
Db    164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy   1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                                     |||      :::||||||   |||
Db    184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy   1196 TTCCAAGCAAGAGACGCTGACATTGAATTT---------------------AGGAAACGG 1234
             :::      :::||||||      |||   :::   |||            |||
Db    204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy   1235 TACTCTACTAAGGGC--------------------------------------------- 1249
                |||
Db    224 ValArgThrSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy   1250 GGTGAAGTTAGATTACAT------------------------TTCCAATTTGAAGGAGGA 1285
          |||      |||::::::                              |||      |||
Db    244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy   1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
             :::   |||         |||||||||         |||
Db    264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy   1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
                      |||      :::   :::                     ::::::|||
Db    283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy   1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
             |||                        :::   ||||||            :::|||
Db    303 ProLys------------------------IleValLeuPro------------GlnArg 310

Qy   1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
                :::::::::::::::   :::|||   :::                  ||||||
Db    311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy   1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
          :::      |||:::   :::      :::                           :::
Db    331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy   1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
                |||            |||            :::||||||   |||
Db    351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy   1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657

Db    368 ------------------------------------------------------------ 368

Qy   1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
                   :::||||||::::::
Db    369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy   1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
                                                                   |||
Db    375 ------------------------------------------------------SerGlu 376

Qy   1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
          |||:::   :::      :::|||   ||||||   |||         :::
Db    377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy   1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTC 1894

Db    394 ------------------------------------------------------------ 394

Qy   1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                |||      |||:::         :::   :::|||:::   ::::::   |||
Db    395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy   1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
          :::
Db    412 Val--------------------------------------------------------- 412

Qy   2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
                      |||      |||:::|||   :::|||   |||:::|||:::|||
Db    413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy   2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
             ||||||                     |||:::      |||   :::|||
Db    429 MetPhePhe---------------------AlaArgProAspSerProPheSer------ 439

Qy   2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                      |||   |||:::               |||         |||||||||
Db    440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy   2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
          |||   ||||||   ||||||   :::::::::         |||   |||   |||:::
Db    456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy   2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
          |||      |||:::         :::   |||      ||||||:::      :::|||
Db    476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy   2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
             |||      |||               ::::::      :::   |||:::
Db    494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy   2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
                   ||||||||||||||||||      :::||||||||||||   :::   |||
Db    514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy   2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
             ||||||||||||||||||||||||:::   |||||||||||||||      |||
Db    534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy   2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
             |||   :::         |||   :::   |||   |||         :::|||
Db    554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573

Qy   2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
          ||||||||||||      |||   :::|||               :::   |||   :::
Db    574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy   2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
          :::   :::|||   ||||||         |||         |||   :::|||||||||
Db    594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy   2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
          |||:::   |||:::                           :::|||
Db    614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy   2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
          :::   ||||||:::   ||||||      ||||||:::      |||
Db    634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653

Qy   2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
          |||:::      :::   ||||||   ||||||||||||||||||   ||||||
Db    654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy   2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
          ||||||   :::||||||::::::||||||      |||:::|||   :::   :::
```

```
Db        673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::|||||||:::   ||||||                      :::      :::   :::
Db        693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy       2927 AGGAAACTCACTGTTGGA 2944
              :::   :::|||:::|||
Db        713 LysAspValThrLeuGly 718

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-507-10

Alignment Scores:
Pred. No.:              2.73e-38      Length:          726
Score:                  463.00        Matches:         207
Percent Similarity:     38.11%        Conservative:    123
Best Local Similarity:  23.90%        Mismatches:      278
Query Match:            6.42%         Indels:          258
DB:                     4             Gaps:            37

US-09-302-812-3 (1-4069) x US-09-511-507-10 (1-726)

Qy        599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
              :::   :::|||      |||              |||         |||      |||
Db         27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              |||      :::                           ||||||:::      :::
Db         47 GluSerLysGluAsp---------------------------ProGluGluProLysSer 57

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
              ::::::   |||  |||:::  |||  |||        ||||||   ||||||
Db         58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy        779 ---------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                   :::|||   :::   ||||||   |||:::|||:::
Db         78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy        815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
              :::|||   :::                |||         ||||||   :::
Db         98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy        863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
              :::|||            ||||||      |||   |||:::
Db        116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy        923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
                         ||||||      ||||||   :::                  ||||||
Db        126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy        983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
                 |||      |||::::::|||:::   :::|||:::
Db        137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy       1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
                    ||||||                         |||   :::      :::
Db        151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy       1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
                 ||||||      ||||||::::::               ||||||      :::
Db        164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy       1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                                     |||      :::||||||   |||
Db        184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy       1196 TTCCAAGCAAGAGACGCTGACATTGAATTT---------------------AGGAAACGG 1234
                 :::      :::||||||   |||                           |||
Db        204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy       1235 TACTCTACTAAGGGC--------------------------------------------- 1249
                    |||
Db        224 ValArgThrSerSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy       1250 GGTGAAGTTAGATTACAT-----------------------TTCCAATTTGAAGGAGGA 1285
              |||      |||::::::                             |||      |||
Db        244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy       1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
                 :::   |||         |||||||||      |||
Db        264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy       1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
                          |||      :::      :::   |||         ::::::|||
Db        283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy       1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
                 |||                        :::   ||||||            :::|||
Db        303 ProLys------------------------IleValLeuPro------------GlnArg 310

Qy       1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
                 ::::::::::::::   :::|||   :::                    ||||||
Db        311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy       1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
                 :::   |||:::      :::   :::                        :::
Db        331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy       1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
                 |||         |||         :::||||||   |||
Db        351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy       1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657

Db        368 ------------------------------------------------------------ 368

Qy       1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
                    :::||||||::::::
Db        369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy       1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
                                                                       |||
Db        375 ------------------------------------------------------SerGlu 376

Qy       1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
              |||:::   :::      :::|||   ||||||   |||         :::
Db        377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy       1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTC 1894

Db        394 ------------------------------------------------------------ 394

Qy       1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
```

```
                    |||        |||:::           :::   :::|||:::   ::::::   |||
Db      395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy     1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
            :::
Db      412 Val--------------------------------------------------------- 412

Qy     2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
                        |||   |||:::|||   :::|||   |||:::|||:::|||
Db      413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy     2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
               ||||||                    |||:::   |||   :::|||
Db      429 MetPhePhe--------------------AlaArgProAspSerProPheSer------ 439

Qy     2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                        |||   |||:::           |||         |||||||||
Db      440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy     2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
            |||   ||||||   ||||||   :::::::::              |||   |||   |||:::
Db      456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy     2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
            |||      |||:::         :::   |||      ||||||:::         :::|||
Db      476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy     2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
               |||      |||               ::::::      :::   |||:::
Db      494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy     2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
                     |||||||||||||||||         :::||||||||||||   :::      |||
Db      514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy     2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
               ||||||||||||||||||||||||:::   ||||||:::::::::      |||
Db      534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy     2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
               |||   :::         |||   :::   |||   |||         :::|||
Db      554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573

Qy     2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
            ||||||||||||      |||   :::|||               :::   |||   :::
Db      574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy     2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
            :::   :::|||   ||||||         |||         |||   :::|||||||||
Db      594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy     2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
            |||:::   |||:::                        :::|||
Db      614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy     2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
            :::   ||||||:::   ||||||      ||||||:::         |||
Db      634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653

Qy     2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
            |||:::      :::   ||||||   ||||||||||||||||||   ||||||
Db      654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy     2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
            ||||||   :::||||||:::::::||||||      |||:::|||   :::   :::
Db      673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692

Qy     2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
            :::||||||:::   ||||||               :::         :::   :::
Db      693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy     2927 AGGAAACTCACTGTTGGA 2944
            :::   :::|||:::|||
Db      713 LysAspValThrLeuGly 718
```

Search completed: May 26, 2004, 17:19:14
Job time : 139.042 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 26, 2004, 17:10:29 ; Search time 169.187 Seconds
(without alignments)
13416.302 Million cell updates/sec

Title: US-09-302-812-3
Perfect score: 7208
Sequence: 1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 2298626

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150056_5800/app_query.fasta_1.12437
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=6333148_@CGN_1_1_79_@runat_26052004_150056_5800
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	72.0	976	9	US-09-973-451-4	Sequence 4, Appli
2	4635.5	64.3	977	9	US-09-973-451-2	Sequence 2, Appli
3	4395	61.0	968	9	US-09-973-451-6	Sequence 6, Appli
4	1081	15.0	768	9	US-09-973-451-8	Sequence 8, Appli
5	595	8.3	546	12	US-10-425-114-60000	Sequence 60000, A
6	463	6.4	726	9	US-09-973-451-10	Sequence 10, Appl
7	217.5	3.0	200	12	US-10-424-599-184988	Sequence 184988,
8	188	2.6	180	12	US-10-424-599-156445	Sequence 156445,
9	186.5	2.6	2476	10	US-09-824-574-7	Sequence 7, Appli
10	179.5	2.5	300	12	US-10-424-599-233915	Sequence 233915,
11	177	2.5	1163	12	US-10-221-278-197	Sequence 197, App
12	177	2.5	1163	15	US-10-291-172-197	Sequence 197, App
13	175	2.4	1111	9	US-09-815-242-12955	Sequence 12955, A
14	174	2.4	1274	12	US-10-282-122A-43837	Sequence 43837, A
15	173.5	2.4	748	9	US-09-864-761-43244	Sequence 43244, A
16	173	2.4	1948	14	US-10-032-585-7611	Sequence 7611, Ap
17	172.5	2.4	498	12	US-10-424-599-196154	Sequence 196154,
18	167	2.3	31	9	US-09-973-451-11	Sequence 11, Appl
19	166	2.3	1163	10	US-09-932-257A-3	Sequence 3, Appli
20	160	2.2	1164	15	US-10-369-493-6564	Sequence 6564, Ap
21	160	2.2	1957	15	US-10-369-493-2070	Sequence 2070, Ap
22	159	2.2	1462	14	US-10-287-218-17	Sequence 17, Appl
23	158.5	2.2	6642	15	US-10-369-493-5013	Sequence 5013, Ap
24	158	2.2	1297	12	US-10-282-122A-71692	Sequence 71692, A
25	158	2.2	1974	15	US-10-369-493-6395	Sequence 6395, Ap
26	156	2.2	1501	9	US-09-924-154-17	Sequence 17, Appl
27	156	2.2	2515	12	US-10-276-774-2651	Sequence 2651, Ap
28	152	2.1	1182	12	US-10-282-122A-53445	Sequence 53445, A
29	152	2.1	5560	12	US-10-263-929-142	Sequence 142, App
30	151.5	2.1	903	12	US-10-282-122A-52328	Sequence 52328, A
31	151.5	2.1	1401	16	US-10-287-226-142	Sequence 142, App
32	150.5	2.1	408	12	US-10-424-599-224071	Sequence 224071,
33	150.5	2.1	1169	12	US-10-282-122A-71179	Sequence 71179, A
34	150.5	2.1	1312	15	US-10-369-493-1950	Sequence 1950, Ap
35	150	2.1	407	12	US-10-282-122A-71177	Sequence 71177, A
36	150	2.1	1905	15	US-10-259-194A-86	Sequence 86, Appl
37	149.5	2.1	691	15	US-10-108-260A-4345	Sequence 4345, Ap
38	149.5	2.1	1210	14	US-10-205-823-264	Sequence 264, App
39	149.5	2.1	1232	14	US-10-116-712-664	Sequence 664, App
40	149.5	2.1	1232	14	US-10-116-712-669	Sequence 669, App
41	149	2.1	26	9	US-09-973-451-20	Sequence 20, Appl
42	149	2.1	29	9	US-09-973-451-12	Sequence 12, Appl
43	148.5	2.1	665	10	US-09-820-843A-107	Sequence 107, App
44	148.5	2.1	677	14	US-10-060-036-4552	Sequence 4552, Ap
45	148.5	2.1	677	15	US-10-295-027-1306	Sequence 1306, Ap

ALIGNMENTS

RESULT 1
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT

```
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-973-451-4

Alignment Scores:
Pred. No.:              0            Length:         976
Score:                  5190.00      Matches:        976
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            72.00%       Indels:         0
DB:                     9            Gaps:           0

US-09-302-812-3 (1-4069) x US-09-973-451-4 (1-976)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy    227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy    287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy    347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy    407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy    467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy    527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy    587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy    647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy    707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy    767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy    827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy    887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy    947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy   1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy   1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy   1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy   1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy   1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy   1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy   1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy   1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy   1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy   1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy   1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy   1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy   1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy   1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy   1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy   1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy   2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy   2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660
```

```
Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

RESULT 2
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-2

Alignment Scores:

Pred. No.:	0	Length:	977
Score:	4635.50	Matches:	874
Percent Similarity:	92.83%	Conservative:	32
Best Local Similarity:	89.55%	Mismatches:	69
Query Match:	64.31%	Indels:	1
DB:	9	Gaps:	1

US-09-302-812-3 (1-4069) x US-09-973-451-2 (1-977)

```
Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              |||:::||||||||||||||||||||||||||||||      ||||||      ||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              ||||||   ||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||   |||||||||   ||||||||||||||||||||||||||||||      |||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        344 GGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              ||||||||||||||||||||||||||||||||||   |||||||||||||||   ||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
              |||||||||:::|||||||||   |||||||||||||||||||      |||||||||
Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||   |||:::|||:::|||||||||||||||||||      ||||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              |||||||||||||||||||||:::|||:::|||      ||||||||||||||||||
Db        181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              :::|||||||||:::|||||||||||||||||||||||||      ||||||||||||||||
Db        201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
```

```
Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db    721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy   2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db    741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy   2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db    761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy   2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db    781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy   2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db    801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy   2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db    821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy   2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db    841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy   2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db    861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy   2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db    881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy   2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db    901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy   2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db    921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy   2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
```

```
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 3
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-973-451-6

Alignment Scores:
Pred. No.:              0              Length:        968
Score:                  4395.00        Matches:       837
Percent Similarity:     90.99%         Conservative:  52
Best Local Similarity:  85.67%         Mismatches:    78
Query Match:            60.97%         Indels:        10
DB:                     9              Gaps:          6

US-09-302-812-3 (1-4069) x US-09-973-451-6 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db        217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db        314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy       1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db        333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy       1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db        353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy       1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db        372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValGluCysArgSer 391

Qy       1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db        392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy       1424 GCAGAGGACAGAAGAAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db        412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy       1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db        432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db        452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db        472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy       1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
```

```
Db        492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy       1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy       1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||||     ||||||:::||||||||||||||||||||||
Db        712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||||||||||:::||||||   :::||||||||||
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              |||||||||||   ||||||   |||||||||||||||||||||||||||||||||||||
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              |||||||||   |||||||||   |||         ||||||:::   ||||||
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 4
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-973-451-8

Alignment Scores:
Pred. No.:              1.92e-77       Length:         768
Score:                  1081.00        Matches:        246
Percent Similarity:     57.89%         Conservative:   99
Best Local Similarity:  41.28%         Mismatches:     197
Query Match:            15.00%         Indels:         54
DB:                     9              Gaps:           16

US-09-302-812-3 (1-4069) x US-09-973-451-8 (1-768)

Qy       1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGAAACCAAACATCAAAGAACA 1471
              |||:::   |||:::   |||                 :::   ||||||:::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
                                        |||   :::|||   |||   |||   |||
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
                 |||      :::|||   :::         |||   |||            |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
              |||   :::                  |||   |||      ::::::   :::
```

```
Db      95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy    1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
Db     115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy    1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
Db     131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy    1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
Db     151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy    1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
Db     171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy    1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
Db     191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy    1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
Db     211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy    2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
Db     231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy    2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
Db     251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy    2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
Db     271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy    2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
Db     291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy    2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
Db     311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy    2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
Db     331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy    2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db     351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy    2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
Db     371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy    2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
Db     391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy    2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
Db     411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy    2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
Db     431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy    2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db     451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy    2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
Db     471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy    2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
Db     491 ProGly-----------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy    2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
Db     503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy    2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
Db     523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy    2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
Db     543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy    2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
Db     562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy    3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
Db     582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy    3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
Db     599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614
```

RESULT 5
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

Alignment Scores:

Pred. No.:	1.5e-38	Length:	546
Score:	595.00	Matches:	151
Percent Similarity:	50.34%	Conservative:	72
Best Local Similarity:	34.09%	Mismatches:	150
Query Match:	8.25%	Indels:	70
DB:	12	Gaps:	12

US-09-302-812-3 (1-4069) x US-10-425-114-60000 (1-546)

```
Qy    1823 TTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCT------------------AAG 1864
Db      88 LeuAlaAspAlaLeuThrAspLeuArgLeuAlaLeuAlaLeuProAlaLeuProProArg 107

Qy    1865 AAATGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCT 1924
```

```
                  |||      |||||      |||:::|||   :::|||     |||:::|||
Db      108 AlaAlaAspGlyLeuAlaLeu------PhePheAspAspLeuLeuSerArgAlaGlnAla 125

Qy     1925 CAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAAT 1984
            :::    :::    ::::::|||::::::  ::::::  |||   |||||
Db      126 ArgGlyTrpPheSerGluValValProAsnLeuAlaArgLeuLeuLeuArgLeuProThr 145

Qy     1985 ATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCC------------ 2032
            :::                                |||   |||
Db      146 LeuLeuGluAspHisTyr------------AlaLysAlaGlyHisGlyAlaSerGlyLeu 161

Qy     2033 ------------------------------ATCACAATGTCGCAGGAACAGATTGCCAGT 2062
                                          :::   :::|||||||||   :::|||:::
Db      162 ArgValMetAlaSerGlnAspAlaGlyValValLeuLeuSerGlnGluLeuValAlaAla 181

Qy     2063 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAG 2122
            ||||||      |||   ||||||   ||||||
Db      182 LeuLeuThrCysAlaLeuPheCysLeuPheProThrAlaGlyArgAlaGln--------- 198

Qy     2123 TATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTT---------------GAGGGA 2167
               :::      |||   |||||||||:::   ||||||
Db      199 ---AlaCysLeuProThrIleAsnPheAspGlyLeuPheAlaProLeuIleHisAsnAla 217

Qy     2168 CGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACA 2227
            |||   |||::::::   :::|||::::::   |||      ||||||   |||||||||
Db      218 ArgGlnSerGlnGlu---GlnLysValArgCysLeuValHisTyrPheGluArgValThr 236

Qy     2228 GAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTT--------------- 2272
            :::   |||||||||||||||:::|||   |||:::   |||
Db      237 AspSerMetProThrGlyLeuValSerPheGluArgLysValLeuProArgArgAlaLeu 256

Qy     2273 ------------------GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACA 2314
                                 |||         |||               ||||||
Db      257 SerAspGlyValProTyrProAspIleHisAlaTrpValAlaSerSerAlaProLeuCys 276

Qy     2315 CGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTG 2374
            :::      |||         |||   ||||||:::      |||      |||||||||
Db      277 GlnPheThrValPheSerSerGlyPheIleGluAspGluGluGlnGluAlaLeuGlnVal 296

Qy     2375 GATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAA 2434
            ||||||||||||:::::::::||||||||      |||   |||   ||||||||||||
Db      297 AspPheAlaAsnLysTyrLeuGlyGlyGlyAlaLeuSerTrpGlyCysValGlnGluGlu 316

Qy     2435 ATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGAT 2494
            |||||||||:::||||||||||||||||||:::        ||||||      ::::::
Db      317 IleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeuSerCysMetGlu 336

Qy     2495 CACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCT 2554
               ||||||   :::   |||   |||   |||::::::|||:::|||   ||||||
Db      337 AspAsnGluAlaIleGluIlePheGlyAlaGluArgPheSerGlnTyrMetGlyTyrGly 356

Qy     2555 GAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGG 2614
               ::::::|||:::         :::   |||               |||      |||
Db      357 SerSerPheArgPheValGlyAspTyrLeuAspThrLysProPheAspSerMetGlyArg 376

Qy     2615 CGCTGCACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTT 2674
            |||   |||   ||||||||||||||||||||||||         |||         :::
Db      377 ArgArgThrArgIleValAlaIleAspAlaLeuAspCysProAlaArgLeu---HisTyr 395

Qy     2675 GTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------ 2728
                     :::   ||||||:::||||||||||:::|||||||||
Db      396 GluSerGlyCysLeuLeuArgGluValAsnLysAlaPheCysGlyPhePheAspGlnSer 415

Qy     2729 ------------------------------------------------------CCTGGA 2734

Db      416 LysHisHisLeuTyrAlaLysLeuPheGlnAspLeuHisAsnLysAspAspPheSerSer 435

Qy     2735 GTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGG 2794
            ::::::|||                     |||:::|||||||||||||||||||||||||||
Db      436 IleAsnSerSerGluTyrValGlyValSerThrGlyAsnTrpGlyCysGlyAlaPheGly 455

Qy     2795 GGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGAT 2854
            |||:::      :::|||::::::||||||   :::||||||:::   |||   |||
Db      456 GlyAsnProGluIleLysSerMetIleGlnTrpIleAlaAlaSerGlnAlaLeuArgPro 475

Qy     2855 GTGGTT---TATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCAC 2911
               |||   |||:::||||||   |||         |||   |||   :::      :::
Db      476 PheValAsnTyrTyrThrPheGluAspValSerLeuGlnArg---LeuGluGluValIle 494

Qy     2912 ATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATAC 2971
               ::::::                     |||||||||:::::::::   :::|||:::   |||
Db      495 GlnTrpIleArgLeuHisGlyTrpThrValGlyGluLeuTrpHisMetLeuMetGluTyr 514

Qy     2972 TACAATGAA 2980
               ::::::
Db      515 SerSerGln 517
```

RESULT 6
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-973-451-10

Alignment Scores:

Pred. No.:	6.41e-28	Length:	726
Score:	463.00	Matches:	207
Percent Similarity:	38.11%	Conservative:	123
Best Local Similarity:	23.90%	Mismatches:	278
Query Match:	6.42%	Indels:	258
DB:	9	Gaps:	37

US-09-302-812-3 (1-4069) x US-09-973-451-10 (1-726)

```
Qy      599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
               :::   :::|||      |||               |||         |||   |||
Db       27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy      659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
               |||      :::                              ||||||:::      :::
Db       47 GluSerLysGluAsp--------------------------ProGluGluProLysSer 57

Qy      719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
               ::::::      |||   |||:::   |||   |||         ||||||   ||||||
Db       58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy      779 --------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                :::|||   :::   ||||||   |||:::|||:::
Db       78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy      815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
            :::|||   :::               |||         ||||||         :::
Db       98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy      863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGAAGACTGTGCAAGTTGTCAGCAA 922
```

```
Db     116 AsnHis------------LysSerThrGluProMetGluGlu----------------- 125

Qy     923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
Db     126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy     983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
Db     137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy    1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
Db     151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy    1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
Db     164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy    1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
Db     184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy    1196 TTCCAAGCAAGAGACGCTGACATTGAATTT---------------------AGGAAACGG 1234
Db     204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy    1235 TACTCTACTAAGGGC--------------------------------------------- 1249
Db     224 ValArgThrSerSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy    1250 GGTGAAGTTAGATTACAT-----------------------TTCCAATTTGAAGGAGGA 1285
Db     244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy    1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
Db     264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy    1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
Db     283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy    1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
Db     303 ProLys------------------------IleValLeuPro------------GlnArg 310

Qy    1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
Db     311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy    1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
Db     331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy    1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
Db     351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy    1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657
Db     368 ------------------------------------------------------------ 368

Qy    1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
Db     369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy    1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
Db     375 ---------------------------------------------------------SerGlu 376

Qy    1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
Db     377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy    1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTC 1894
Db     394 ------------------------------------------------------------ 394

Qy    1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
Db     395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy    1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
Db     412 Val--------------------------------------------------------- 412

Qy    2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
Db     413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy    2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
Db     429 MetPhePhe--------------------AlaArgProAspSerProPheSer------ 439

Qy    2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
Db     440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy    2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
Db     456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy    2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
Db     476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy    2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
Db     494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy    2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
Db     514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy    2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
Db     534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy    2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
Db     554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573

Qy    2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
Db     574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy    2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
Db     594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy    2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db     614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy    2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
Db     634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653

Qy    2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
Db     654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy    2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
Db     673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692
```

```
Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::|||||:::   ||||||                :::          :::   :::
Db        693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy       2927 AGGAAACTCACTGTTGGA 2944
              :::   :::|||:::|||
Db        713 LysAspValThrLeuGly 718

RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(200)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

Alignment Scores:
Pred. No.:              1.8e-08        Length:         200
Score:                  217.50         Matches:        53
Percent Similarity:     45.20%         Conservative:   27
Best Local Similarity:  29.94%         Mismatches:     68
Query Match:            3.02%          Indels:         29
DB:                     12             Gaps:           4

US-09-302-812-3 (1-4069) x US-10-424-599-184988 (1-200)

Qy       2039 ATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGA 2098
              ::::::||||||      |||          :::          ||||||   ||||||
Db         33 LeuThrGlnGluThrSerAlaProProProSerArgThrLeuPheCysLeuPhePro--- 51

Qy       2099 CGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCA--------------------- 2137
                                             |||    |||
Db         52 ------------------------ValSerAspArgProGlyIleHisLeuProMetIle 63

Qy       2138 ------GACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAA 2191
                    |||::::::         |||:::|||      |||    :::
Db         64 AsnSerAspValSerSerGlySerLeuTyrGluAspTyrSerGlnLysProGluSerThr 83

Qy       2192 CTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTG 2251
                    :::      |||   :::|||::::::   :::   |||   |||:::|||
Db         84 ThrGlyArgIleAlaHisTyrGlyGlnArgIleSerSerGluMetProLysGlyIleVal 103

Qy       2252 ACATTTACAAGACAGAGTCTTGAA-----------------------GATTTCCAGAA 2287
              :::|||   |||:::   |||                           :::|||:::
Db        104 SerPheGluArgLysValLeuProPheLysAsnAspSerIleHisIleSerTyrProAsp 123

Qy       2288 ---------TGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGT 2338
                       |||                ||||||   |||   |||            |||
Db        124 AlaAsnPheTrpSerThrSerAlaIleProLeuCysArgPheGluValHisSerSerGly 143

Qy       2339 ACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGA 2398
                 ||||||:::         |||         :::||||||   ||||||:::::::::|||
Db        144 LeuIleGluAspGlnSerSerGlyAlaAlaGluValAspLeuAlaAsnLysTyrLeuGly 163

Qy       2399 GGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTAATCAATCCTGAG 2458
              ||||||                  |||   |||||||||||||||||||:::::::::||||||
Db        164 GlyGlyAlaLeuGlyArgGlyCysValGlnGluGluIleArgPheMetValSerProGlu 183

Qy       2459 TTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTA 2509
              |||               ||||||          :::      ||||||   :::
Db        184 LeuGluAlaGlyMetLeuPheLeuProAlaMetAlaAspAsnGluAlaIle 200

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

Alignment Scores:
Pred. No.:              3.99e-06       Length:         180
Score:                  188.00         Matches:        49
Percent Similarity:     44.17%         Conservative:   23
Best Local Similarity:  30.06%         Mismatches:     41
Query Match:            2.61%          Indels:         50
DB:                     12             Gaps:           4

US-09-302-812-3 (1-4069) x US-10-424-599-156445 (1-180)

Qy       2696 GAGCTGAACAAGGCTTACTGTGGATTTCTC---------------CGTCCT--------- 2731
              |||:::||||||||:::||||||||||||                  :::|||
Db          7 GluIleAsnLysAlaPheCysGlyPheLeuTyrGlnCysLysTyrGlnProTyrGlnLys 26

Qy       2731 ------------------------------------------------------------ 2731

Db         27 IleLeuGlnGluAsnGlyCysThrSerAlaLeuPheTyrAlaAlaThrSerThrSerMet 46

Qy       2731 ------------------------------------------------------------ 2731

Db         47 GluThrAspGluGlyGluIleSerAsnHisLysIleThrAsnSerGlnAsnAspTyrHis 66

Qy       2732 GGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTT 2791
              |||:::            |||         ||||||||||||||||||||||||||||||
Db         67 GlyMetAspGlnGlyAsnAsnIleGlyValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 86

Qy       2792 GGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGA 2851
              |||||||||         :::|||   :::||||||   |||||||||:::   |||   |||
Db         87 GlyGlyAspProGluValLysThrIleIleGlnTrpLeuAlaAlaSerGlnAlaLeuArg 106

Qy       2852 GAT---GTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
              :::      |||:::|||||||||          |||      ::::::   :::      :::
Db        107 ProPheIleAlaTyrTyrThrPheGly---LeuGluAlaLeuGlnSerLeuAspGluVal 125

Qy       2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
                 ::::::         ::::::   ||||||||||||::::::      :::|||:::
Db        126 AlaHisTrpIleLeuSerGlnArgTrpThrValGlyAspLeuTrpAsnMetLeuIleGlu 145

Qy       2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCA 3028
              |||               :::         :::         |||               |||   |||
Db        146 TyrSerIleAsnArgSerLysGlyGluThrAsnValGlyPheLeuGlnTrpLeuLeuPro 165
```

```
Qy       3029 TTCATATAC 3037
                 ||||||
Db        166 SerIleTyr 168
```

RESULT 9
US-09-824-574-7
; Sequence 7, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jnne, Olli A,
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 2476
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-824-574-7

```
Alignment Scores:
Pred. No.:               1.54e-05      Length:          2476
Score:                   186.50        Matches:         216
Percent Similarity:      35.48%        Conservative:    142
Best Local Similarity:   21.41%        Mismatches:      379
Query Match:             2.59%         Indels:          273
DB:                      10            Gaps:            53

US-09-302-812-3 (1-4069) x US-09-824-574-7 (1-2476)

Qy         67 AGGCAGGCGCTGAGAAGGACGCGCAGTCCATCTCTCTCAGGTTAGTGAAATGAGGCTCTC 126
              |||        :::    ::::::|||   |||   |||          |||   |||
Db        967 ArgProGlyThrLysGlyLysLysAlaProAspLeuLysGly---------GluThrLeu 983

Qy        127 CGC-----------------GGGGCCGGCCCGGGGACAGTGCGCTGCTGGTCCCAGCAT 168
              :::                 :::       ||||||   |||       :::
Db        984 LysArgGluGlnGluTrpAspSerSerSerAspGlyThrGluArgLeuProGluGluGlu 1003

Qy        169 GAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGC--------------------GAC 207
              |||   ||||||           :::    ||||||                     :::
Db       1004 GluIleGlyProPheSerLysGlyIleLysGlnSerLysThrAspThrAlaGlyGlyGlu 1023

Qy        208 CCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAG 267
                 |||:::::::::           |||       :::   |||||||||:::
Db       1024 LysLysGlyLysLysTrpLysAspLysSerCysGluLysLysGluGluLeuSerAspSer 1043

Qy        268 GCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTC 327
                 :::   |||     :::|||   |||   |||   |||       :::
Db       1044 ValAspLysLeuPro---GlyLysGlyAspSer--CysAspSerSerGluAspLysLysT 1062

Qy        328 GCCAG----------------------------------------CCTGCG---TCCCAGG 345
                 :::                                       ||||||   |||   |
Db       1062 hrArgAsnArgValSerLeuArgGluLysLysArgPheSerLeuProAlaLysSerProG 1082

Qy        346 GCAGGCGG--------------------------------------------------G 354
              ||:::|||                                                  |
Db       1082 lyLysArgProGluCysSerSerSerAspThrGluLysSerLeuLysGlyGlnCysCysA 1102

Qy        355 ACAGCACAGAGGCAGCGCCACCTCGCTTGTT----TTCAAACAAAAGACTATTACCAGTT 410
              ||||||||||         |||    |||:::        :::::::::       :::|||
Db       1102 spSerThrGluLysArgProLysArgIleAspLeuArgGluArgArgAsnSerSerSerL 1122

Qy        411 GGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGT------TTGGATAGTAAAG 464
                 :::||||||    :::|||:::|||    |||    |||                |
Db       1122 ysArgAsnThrLysGluValLysSerAlaSerSerSerSerAspAlaGluGlySerSerG 1142

Qy        465 AAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACC 524
              ||:::|||       :::   :::               |||:::   :::|||         :
Db       1142 luAspAsnLysLysGlnLysLysGlnArgThrSerAlaLysLysLysThrGlyAsnThrL 1162

Qy        525 AACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTG 584
              ::                     :::   :::                     :::|||
Db       1162 ysGluLysLysArgAsnSerLeuArgAlaThrProLysArgLysGlnValAspIleThrS 1182

Qy        585 AAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATG 644
                    |||:::      :::      |||   :::|||      :::      :::|
Db       1182 erSerSerSerAspIleGlyAspAspAspGlnAsnSerAlaGlyGluGluSerSerAspG 1202

Qy        645 AAGGGAAA----------------------------------CACACG---------- 658
              ||  |||                                   ||||||
Db       1202 luGlnLysIleLysProValThrGluAsnLeuValLeuProSerHisThrGlyPheCysG 1222

Qy        659 --------------GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAG 704
                            |||   |||   :::|||   |||   ||||||
Db       1222 lnSerSerGlyAspGluAlaLeuSerLysSerValProAlaThrVal------------ 1237

Qy        705 AGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAG 764
                                                       ||||||   |||   :
Db       1238 ----------------------------------------AspAspAspAspAspAspA 1244

Qy        765 ATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAA 824
              ::      |||||||||      ::::::   |||               ||||||
Db       1244 snAspProGluAsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuS 1264

Qy        825 AGCAG-----------ACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGA 872
                             :::::::::||||||          |||
Db       1264 erSerAspGluAspGlySerSerAspAspGluProAspGlyGlyGlyLysLysArgIleG 1284

Qy        873 AGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAG 932
                          |||   :::|||                  :::::::::::::::
Db       1284 lyLysGlnSerGluGluSerProAlaAspAsp---GlyGluLeuArgArgGluGlnLeuA 1303

Qy        933 ACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAA 992
              |||         :::         ||||||   ||||||:::                |||
Db       1303 laValAsnGlnValAsnSerGluSerAspSerAspSerGluGlu------------SerL 1319

Qy        993 AAAATGAC---------AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTC 1043
              ||            ::::::|||:::|||   :::         |||   :::
Db       1319 ysLysProArgTyrArgHisArgLeuLeuArgHisLysLeuThrLeuSerAspGlyGluS 1339

Qy       1044 CA---TTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATA 1100
                   ||||||   ::::::|||:::          :::
Db       1339 erGlyGluGluLysProThrLysProLysGluHisLysGluAlaLysGlyArgAsnArgA 1359

Qy       1101 GTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATG 1160
                       ||||||      |||:::|||            |||   ||||||   |
Db       1359 rgLysValSerSerGluAspSerGluAspThr------------AspPheGlnGluSerG 1375

Qy       1161 GT---------------AGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAA 1205
              ||               |||   |||:::         :::|||         |||:
Db       1375 lyValSerGluGluValSerGluSerGluAspGluGlnArgProArgThrArgSerAlaL 1395

Qy       1206 GAGACGCTGACATTGAATTTAGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGAT 1262
              ::   |||::::::|||      :::|||   |||      |||         |||:
Db       1395 ysLysAlaGluLeuGluGluAsnGlnArgSerTyrLysGlnLysLysLysArgArgArgI 1415

Qy       1263 TACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAAC 1322
              ::      |||   |||         |||            :::            |||
Db       1415 leLysValGln---GluAspSerSerSerGluAsnLysSerHisSerGluGluAspLysL 1434

Qy       1323 TACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGA--------------------A 1361
                    |||:::               |||                              :
Db       1434 ysGluGlyAspGluGluAspGluAspGluAspGluGluAspGluGluAspGluAsnAspA 1454

Qy       1362 ATTCTAAGCAACATGGAAAA--------------------------AAGGATTCTAAAA 1394
              ::||||||      ||||||                          ||||||   |||:
Db       1454 spSerLysSerProGlyLysGlyArgLysLysIleArgLysIleLeuLysAspAspLysL 1474

Qy       1395 TC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGG 1451
              ::   |||:::               |||   :::   |||:::|||||||||       |
```

```
Db      1474 euArgThrGluThrGlnAsnAlaLeuLysGluGluGluArgArgLysArgIleAlaG 1494

Qy      1452 AAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA---CCTCACCTTT 1508
             ||   :::   :::|||        |||:::   :::   :::          |
Db      1494 luArgGluArgGluArg------GluLysLeuArgGluValIleGluIleGluAspAlaS 1512

Qy      1509 CTCCAGATAAGAAGTGGCTTGGAACTCCCATT---------------GAGGAGATGAGAA 1553
             |||||  |||        :::    |||   :::            ||||||    :::
Db      1512 erProThrLysCysProIleThrThrLysLeuValLeuAspGluAsnGluGluThrLysG 1532

Qy      1554 GAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAA 1613
                                        ||||||::::::         |||       :::
Db      1532 lu------------------------ProLeuValGlnValHisArgAsnMetValIleL 1544

Qy      1614 CTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATA 1673
              :::                                       ||||||       |||
Db      1544 ysLeu---------------------------------LysPro---------HisGlnV 1550

Qy      1674 AAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAA 1721
               |||          :::||||||                     |||       :::
Db      1550 alAspGlyValGlnPheMetTrpAsp---------------------CysCysCysGluS 1563

Qy      1722 ATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGAGCTCA 1781
             :::::               |||      :::::::::    ||||||        |||
Db      1563 erVal---------------GluLysThrLysLysSerProGlySerGlyCysIleLeuA 1578

Qy      1782 TTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGA 1841
                         |||    |||                        |||:::
Db      1578 laHisCysMetGlyLeuGlyLys---------------ThrLeuGlnValValSerPheL 1593

Qy      1842 AATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTGATCGATTTC---- 1894
               :::     |||                |||   ||||||    |||||||||:::
Db      1593 euHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeuValValCysProL 1613

Qy      1895 -----------------TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATC 1937
                              |||            |||:::                  :::|
Db      1613 euAsnThrAlaLeuAsnTrpMetAsnGluPheGluLys------------------TrpG 1627

Qy      1938 AGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGC 1997
             ||        |||   |||     |||:::          :::        :::   |||
Db      1627 lnGluGlyLeuAsnAspAsnGluLysLeuGlu------ValSerGluLeuAlaThrValL 1645

Qy      1998 CAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTG 2057
                 |||          |||:::   ::::::   :::            ||||||
Db      1645 ysArgPro---------GlnGluArgSerTyrMetLeuGlnArgTrpGlnGluAspGlyG 1662

Qy      2058 CCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAAT 2117
                 ::::::
Db      1662 lyValMetIleIleGly------------------------------------------- 1667

Qy      2118 CGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAA 2177
                   |||    |||                        |||   :::||||||:::
Db      1668 -----TyrGluMetTyr---------------ArgAsnLeuAlaGlnGlyArgAsnVal- 1680

Qy      2178 GGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAAC 2237
               |||       |||||||||    :::|||         :::   :::   :::      |
Db      1681 --LysSerArgLysLeuLysAspIlePhe------AsnLysAlaLeuValAspProGlyP 1698

Qy      2238 CTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGAT 2297
              ||      :::|||                   |||::::::        |||
Db      1698 roAspPheValValCysAspGluGlyHisIleLeuLysAsn------GluAlaSerAlaV 1716

Qy      2298 GTGAAAAACCCTTG-------------ACACGATTGCATGTCACTTACGAAGGTACC---- 2341
                   |||    :::            :::|||       :::          ||||||
Db      1716 alSerLysAlaMetAsnSerIleLysSerArgArgArgIleIleLeuThrGlyThrProL 1736

Qy      2342 -----------------------------------------ATAGAAGAAAATGGCCAAG 2360
                                                      |||:::||||||      |
Db      1736 euGlnAsnAsnLeuIleGluTyrHisCysMetValAsnPheIleLysGluAsnLeuLeuG 1756

Qy      2361 GCATGCTACAGGTGGATTTTGCAAATCGTTTTGTT------------GGAGGTGGTGTAA 2408
              ||    ::::::   :::|||   ||||||||:::               |||
Db      1756 lySerIleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGlyGlnCysA 1775

Qy      2409 CCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTT 2468
                :::    :::|||           ::::::|||      :::              |||:::
Db      1775 laAspSerThrMetVal------AspValArgValMetLysLysArgAlaHisIleLeuT 1793

Qy      2469 CACGGCTC------------------------------------TTCACTGAGGTGCTG- 2491
                :::                                       |||:::   |||
Db      1793 yrGluMetLeuAlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLysPheLeuP 1813

Qy      2492 -----GATCACAATGAATGTCTAATTATCACAGGTACTGAG---CAGTACAGTGAATACA 2543
                       |||       |||   :::      |||      |||         |||
Db      1813 roProLysHisGluTyrValLeuAlaValArgMetThrAlaIleGlnCysLysLeuTyrG 1833

Qy      2544 CAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACG 2603
                |||  |||   |||:::   |||   |||   :::      |||   :::|||
Db      1833 lnTyrTyr-LeuAspHisLeuThrGlyValGlyAsnSerThrGluGlyGlyArgGly--- 1851

Qy      2604 ACTGCGAGCGGCGCTGCACTG 2624
                :::||||||   |||
Db      1852 ---LysAlaGlyAlaLysLeu 1857
```

RESULT 10
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233915
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_5324C.1.pep
US-10-424-599-233915

Alignment Scores:

Pred. No.:	2.36e-05	Length:	300
Score:	179.50	Matches:	56
Percent Similarity:	34.24%	Conservative:	45
Best Local Similarity:	18.98%	Mismatches:	65
Query Match:	2.49%	Indels:	129
DB:	12	Gaps:	5

US-09-302-812-3 (1-4069) x US-10-424-599-233915 (1-300)

```
Qy      1892 TTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCT 1951
             |||:::|||:::|||:::               |||:::         :::|||   :::||||||
Db        76 PhePheAspGluValMetSerGlyGluGluSerSerLysTrpPheGlnGluValLeuPro 95

Qy      1952 GATATGGTGAAAATTGCACTCTGTCTGCCAAATATT------------------------ 1987
               :::      :::   |||   ||||||::::::
Db        96 AlaLeuGlyAsnLeuLeuLeuArgLeuProSerLeuLeuGluSerHisTyrGlnAsnThr 115

Qy      1988 -----------------------------------TGCACCCAGCCAATACCACTCCTG 2011
                                                |||      :::   ||||||
Db       116 AspAsnMetAlaIleAspGlyGluAlaGlyAlaMetLeuThrThrAlaLeuArgLeuLeu 135

Qy      2012 AAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCT 2071
                :::          :::   ::::::||||||   ||||||:::||||||:::
Db       136 AspSerGlnGlnProGlyIleValPheLeuThrGlnGluLeuIleAlaAlaLeuLeuSer 155

Qy      2072 AATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGT 2131
               :::   ||||||   ||||||                              |||
Db       156 CysSerLeuPheCysLeuPhePro---------------------------ValSerAsp 166
```

```
Qy       2132 TACCCA---------------------------GACATTAACTTCAATCGATTGTTTGAG 2164
                 |||                           |||:::   |||      |||::::::
Db        167 ArgProValIleHisLeuProMetIleAsnPheAspValLeuPheGlySerLeuTyrAsp 186

Qy       2165 GGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC 2224
                    |||   ::::::      |||:::     :::      ||||||:::|||:::
Db        187 AspTyrSerGlnLysGlnGluAsnLysIleTrpCysIleValHisTyrPheGlnArgIle 206

Qy       2225 ACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCA 2284
              :::   :::   |||   |||:::
Db        207 SerSerGluMetProLysGlyIleAla--------------------------------- 215

Qy       2285 GAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATA 2344

Db        215 ------------------------------------------------------------ 215

Qy       2345 GAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGT 2404

Db        215 ------------------------------------------------------------ 215

Qy       2405 GTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATT 2464

Db        215 ------------------------------------------------------------ 215

Qy       2465 ATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACT 2524
                                             :::||||||   :::   |||   |||
Db        216 ------------------------------TyrAsnGluArgIleGluIleValGlyVal 225

Qy       2525 GAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAA 2584
              |||::::::|||   ||||||   :::|||   ::::::|||:::||||||:::
Db        226 GluArgPheSerGlyTyrThrAspHisAlaSerSerPheArgPheSerArgAlaLysAla 245

Qy       2585 GATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGATGCT 2644
              :::|||   |||   |||         ||||||   |||      ::::::
Db        246 GluGlyArgGluGluAspProValGlyArgArgLysThrSerAspLeuSer--------- 262

Qy       2645 CTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATG 2689
                                      :::::::::   |||   ||||||
Db        263 ---------------------GluLysTyrPheProProLysMet 270
```

RESULT 11
US-10-221-278-197
; Sequence 197, Application US/10221278
; Publication No. US20040034208A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/221,278
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
;   LENGTH: 1163
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-221-278-197

```
Alignment Scores:
Pred. No.:              6.52e-05      Length:        1163
Score:                  177.00        Matches:       212
Percent Similarity:     32.78%        Conservative:  144
Best Local Similarity:  19.52%        Mismatches:    405
Query Match:            2.46%         Indels:        325
DB:                     12            Gaps:          49

US-09-302-812-3 (1-4069) x US-10-221-278-197 (1-1163)

Qy        110 AGTGAAATGAGGCTCTCCGCGGGGCCGGCCCGGGGACAGTGCGCTGCTGGT--------- 160
              |||      |||   |||||||||   :::   |||      ::::::|||
Db        132 SerSerGlnArgThrSerAlaGlySerSerSerGlyThrAsnSerSerGlyGlnArgHis 151

Qy        160 ------------------------------------------------------------ 160

Db        152 AspArgGluSerTyrAsnAsnSerGlySerSerSerArgLysLysGlyGlnHisGlySer 171

Qy        161 ---------CCCAGCATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGC 211
                       |||   ::::::   ||||||                     |||
Db        172 GluHisSerLysSerArgSerSerSerProGly---------------LysProGlnAla 186

Qy        212 TGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAG 271
                 :::      :::|||   :::   |||                  |||::::::
Db        187 ValSerSerLeuAsnSerSerHisSerArgSerHisGlyAsnAspHisHisSerLysGlu 206

Qy        272 ---AGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCG 328
                 :::|||         |||:::|||            :::               |||
Db        207 HisGlnArgSerLysSerProArgAspProAspAlaAsnTrp------------AspSer 222

Qy        329 CCAGCCTGCGTCCCAGGGCAGGCGGGACAGCAC--------------------------- 361
              |||:::   ||||||      :::|||||||||
Db        223 ProSerArgValProPheSerSerGlyGlnHisSerThrGlnSerPheProProSerLeu 242

Qy        362 AGAGGCAGCGCCACCTCGCTTGTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT 421
                 :::      |||::::::   |||                     ||||||
Db        243 MetSerLysSerAsnSerMetLeuGlnLysProThrAlaTyrValArgProMetAsp--- 261

Qy        422 AAAGGAATCAAGACAGCGGAA------------TCAGAAAGTTTGGATAGTAAAGAAAAC 469
                 |||   ::::::   |||            ||||||      |||:::   :::
Db        262 ---GlyGlnGluSerMetGluProLysLeuSerSerGluHisTyrSerSerGlnSerHis 280

Qy        470 AACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACAT 529
                 |||:::      |||      ||||||                     :::
Db        281 GlyAsnSerMetThrGluLeuLysProSerSer---------------------LysAla 293

Qy        530 AATGTAGAAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCA------------ 577
              ::::::   ||||||      ||||||      ||||||   |||
Db        294 HisLeuThrLysLeuLysIleProSerGln---ProLeuAspAlaSerAlaSerGlyAsp 312

Qy        578 ------------------CTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACT 619
                                 :::   ::::::               |||
Db        313 ValSerCysValAspGluIleLeuLysGluMetThrHisSerTrpProProProLeuThr 332

Qy        620 GCA------GCAATGTGTAAGTGGCAAAATGAA---------------GGGAAACACACG 658
              |||         ||||||   :::                       |||   :::
Db        333 AlaIleHisThrProCysLysThrGluProSerLysPheProPheProThrLysGluSer 352

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              :::|||            :::                  |||   :::   ||||||
Db        353 GlnGlnSerAsnPheGlyThrGlyGluGlnLysArgTyrAsnProSerLysThrSerAsn 372

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGAT---------------CACAGTGACACA 763
                 :::      :::|||         ||||||                  |||:::
Db        373 GlyHisGlnSerLysSerMetLeuLysAspAspLeuLysLeuSerSerSerGluAspSer 392

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              |||   |||::::::   |||                  |||         :::
Db        393 AspGlyGluGlnAspCysAspLys------------ThrMetProArgSerThrProGly 408

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
                 :::         |||   |||         :::   :::
Db        409 SerAsnSerGluProSerHisHisAsnSerGluGlyAlaAspAsnSerArgAspAspSer 428

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              |||   |||   |||   :::   :::|||
```

```
Db    429 SerSerHisSerGlySerGluSerSerSer------------------------------ 438
Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    439 ---------------------GlySerAspSerGluSerGluSerSerSerSerAspSer 451
Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    452 Glu------------------------AlaAsnGluProSerGlnSerAlaSerProGlu 463
Qy   1064 CCCGAATCACCG------------ATGGATGTGGATAAT--------------------- 1090
Db    464 ProGluProProProThrAsnLysTrpGlnLeuAspAsnTrpLeuAsnLysValAsnPro 483
Qy   1090 ------------------------------------------------------------ 1090
Db    484 HisLysValSerProAlaSerSerValAspSerAsnIleProSerSerGlnGlyTyrLys 503
Qy   1091 ---------------------TCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAG 1129
Db    504 LysGluGlyArgGluGlnGlyThrGlyAsnSerTyrThrAspThrSerGlyProLysGlu 523
Qy   1130 ACAAGT---------CCAGGTTTTGATGAACAA------GAAGATGGTAGTTCCTCCCAA 1174
Db    524 ThrSerSerAlaThrProGlyArgAspSerLysThrIleGlnLysGlySerGluSerGly 543
Qy   1175 ACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAA--- 1231
Db    544 ArgGlyArgGlnLysSerProAlaGlnSer---AspSerThrThrGlnArgArgThrVal 562
Qy   1232 ---CGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGA--- 1285
Db    563 GlyLysLysGlnProLysLysAlaGluLysAlaAlaAlaGluGluProArgGlyGlyLeu 582
Qy   1286 ------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCT 1339
Db    583 LysIleGluSerGluThrProVal---AspLeuAlaSerSerMetProSerSerArgHis 601
Qy   1340 AGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAAATCACA 1399
Db    602 LysAlaAlaThrLys---GlySerArgLysProAsnIleLysLysGluSerLysSerSer 620
Qy   1400 GATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAA 1459
Db    621 ProArg------------ProThrAlaGlu---LysLysLysTyrLysSerThrSerLys 635
Qy   1460 CATCAAAGAACAGAAAGGAAGATC------------------------------------ 1483
Db    636 SerSerGlnLysSerArgGluIleIleGluThrAspThrSerSerSerAspSerAspGlu 655
Qy   1484 ------------------------------CCTAAATACGTTCCACCTCACCTTTCTCCA 1513
Db    656 SerGluSerLeuProProSerSerGlnThrProLysTyrProGluSerAsnArgThrPro 675
Qy   1514 GATAAG------------------------------------AAGTGGCTTGGAACTCCC 1537
Db    676 ValLysProSerSerValGluGluGluAspSerPhePheArgGlnArgMetPheSerPro 695
Qy   1538 ATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
Db    696 MetGluGluLysGluLeuLeuSer------------------ProLeuSerGluProAsp 709
Qy   1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCT---AAA 1654
Db    710 AspArgTyrProLeuIleValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLys 729
Qy   1655 CCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCA 1714
Db    730 Pro------------TyrLysGluThrGluProProLysGlyGluLysLysAsnValPro 745
Qy   1715 GAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGG 1774
Db    746 GluLysHisThrArgGluAlaGlnLysGlnAlaSerGluLysValSer------------ 761

Qy   1775 GAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
Db    762 ------------------------AsnLysGlyLysArgLysHisLysAsnGluAsp--- 772
Qy   1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTC 1894
Db    773 ------AspAsnArgAlaSerGluSerLysLys--------------------------- 781
Qy   1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTA------------TATCAGTCC 1942
Db    782 ---ProLysThrGluAspLysAsnSerAlaGlyHisLysProSerSerAsnArgGluSer 800
Qy   1943 ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 2002
Db    801 SerLysGlnSerAlaAlaLysGluLysAspLeuLeuProSer---ProAlaGlyProVal 819
Qy   2003 CCACTCCTGAAACAGAAGATGAATCATTCCATC---------ACAATGTCGCAGGAACAG 2053
Db    820 ProSerLysAspProLysThrGluHisGlySerArgLysArgThrIleSerGlnSerSer 839
Qy   2054 ATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATG 2113
Db    840 SerLeuLysSerSerSerAsnSer------------------------AsnLysGluThr 851
Qy   2114 AAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCA 2173
Db    852 SerGlySerSerLysAsnSerSerSerThrSerLysGlnLysLysThrGluGlyLysThr 871
Qy   2174 TCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAA 2233
Db    872 SerSerSerSerLysGluValLys------------------------------GluLys 881
Qy   2234 AAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAA 2293
Db    882 AlaProSer------------------------SerSerSerAsnCysProPro------ 891
Qy   2294 AGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAAT 2353
Db    892 ------SerAlaPro------------------------------ThrLeuAspSerSer 899
Qy   2354 GGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGT 2413
Db    900 LysProArgArgThrLysLeuValPheAspAspArg------------AsnTyrSerAla 915
Qy   2414 GCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGG 2473
Db    916 AspHisTyrLeuGlnGluAlaLysLysLeuLysHisAsnAlaAspAlaLeuSerAspArg 935
Qy   2474 CTC------------------TTCACTGAGGTGCTGGATCACAATGAATGT--------- 2506
Db    936 PheGluLysAlaValTyrTyrLeuAspAlaValValSerPheIleGluCysGlyAsnAla 955
Qy   2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACA------ 2560
Db    956 LeuGluLysAsnAlaGlnGluSerLysSerProPheProMetTyrSerGluThrValAsp 975
Qy   2561 ---------TATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAG 2611
Db    976 LeuIleLysTyrThrMetLysLeuLysAsnTyrLeuAlaProAspAlaThrAlaAlaAsp 995
Qy   2612 CGGCGCTGCACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAG 2671
Db    996 LysArgLeuThrValLeu--CysLeuArgCysGluSerLeuLeuTyrLeuArgLeuPheL 1015
Qy   2672 TTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT--- 2728
Db   1015 ysLeuLysLysGluAsnAlaLeuLys-TyrSerLysThrLeuThrGluHisLeuLysAsn 1034
Qy   2729 ---------------------------CCTGGAGTTTCTTCAGAG--------------- 2746
Db   1035 SerTyrAsnAsnSerGlnAlaProSerProGlyLeuGlySerLysAlaValGlyMetPro 1054
```

```
Qy       2747 ---------------AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTT 2791
                             ||||||          :::||||||:::       ||||||
Db       1055 SerProValSerProLysLeuSerProGlyAsnSerGlyAsnTyrSerSerGlyAlaSer 1074

Qy       2792 GGGGGTGATGCC 2803
                       |||
Db       1075 SerAlaSerAla 1078

RESULT 12
US-10-291-172-197
; Sequence 197, Application US/10291172
; Publication No. US20030228584A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-045
;  CURRENT APPLICATION NUMBER: US/10/291,172
;  CURRENT FILING DATE:  2000-11-08
;  PRIOR APPLICATION NUMBER: 09/693,267
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 09/665,363
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: 09/616,847
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 09/596,193
;  PRIOR FILING DATE: 2000-06-17
;  PRIOR APPLICATION NUMBER: 09/574,454
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: 09/519,705
;  PRIOR FILING DATE: 2000-03-07
;  NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
;   LENGTH: 1163
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-291-172-197

Alignment Scores:
Pred. No.:              6.52e-05       Length:          1163
Score:                  177.00         Matches:         212
Percent Similarity:     32.78%         Conservative:    144
Best Local Similarity:  19.52%         Mismatches:      405
Query Match:            2.46%          Indels:          325
DB:                     15             Gaps:            49

US-09-302-812-3 (1-4069) x US-10-291-172-197 (1-1163)

Qy        110 AGTGAAATGAGGCTCTCCGCGGGGCCGGCCCGGGGACAGTGCGCTGCTGGT--------- 160
              |||        |||    |||||||||    :::      |||        ::::::|||
Db        132 SerSerGlnArgThrSerAlaGlySerSerSerGlyThrAsnSerSerGlyGlnArgHis 151

Qy        160 ------------------------------------------------------------ 160

Db        152 AspArgGluSerTyrAsnAsnSerGlySerSerSerArgLysLysGlyGlnHisGlySer 171

Qy        161 ---------CCCAGCATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGC 211
                          |||    ::::::    ||||||                   |||
Db        172 GluHisSerLysSerArgSerSerSerProGly---------------LysProGlnAla 186

Qy        212 TGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAG 271
                       :::       :::|||    :::    |||                  |||::::::
Db        187 ValSerSerLeuAsnSerSerHisSerArgSerHisGlyAsnAspHisHisSerLysGlu 206

Qy        272 ---AGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCG 328
                 :::|||          |||:::|||              :::            |||
Db        207 HisGlnArgSerLysSerProArgAspProAspAlaAsnTrp------------AspSer 222

Qy        329 CCAGCCTGCGTCCCAGGGCAGGCGGGACAGCAC--------------------------- 361
              |||:::    ||||||        :::|||||||||
Db        223 ProSerArgValProPheSerSerGlyGlnHisSerThrGlnSerPheProProSerLeu 242

Qy        362 AGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT 421
                         :::    |||::::::     |||                     ||||||
Db        243 MetSerLysSerAsnSerMetLeuGlnLysProThrAlaTyrValArgProMetAsp--- 261

Qy        422 AAAGGAATCAAGACAGCGGAA------------TCAGAAAGTTTGGATAGTAAAGAAAAC 469
                 |||    ::::::    |||              ||||||          |||:::   :::
Db        262 ---GlyGlnGluSerMetGluProLysLeuSerSerGluHisTyrSerSerGlnSerHis 280

Qy        470 AACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACAT 529
                 |||:::        |||             ||||||                 :::
Db        281 GlyAsnSerMetThrGluLeuLysProSerSer---------------------LysAla 293

Qy        530 AATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCA------------ 577
              ::::::    ||||||          ||||||       ||||||    |||
Db        294 HisLeuThrLysLeuLysIleProSerGln---ProLeuAspAlaSerAlaSerGlyAsp 312

Qy        578 ------------------CTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACT 619
                                 :::    ::::::    |||                   |||
Db        313 ValSerCysValAspGluIleLeuLysGluMetThrHisSerTrpProProProLeuThr 332

Qy        620 GCA------GCAATGTGTAAGTGGCAAAATGAA---------------GGGAAACACACG 658
              |||                ||||||    :::                      |||    :::
Db        333 AlaIleHisThrProCysLysThrGluProSerLysPheProPheProThrLysGluSer 352

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              :::|||                 :::                      |||    :::    ||||||
Db        353 GlnGlnSerAsnPheGlyThrGlyGluGlnLysArgTyrAsnProSerLysThrSerAsn 372

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGAT---------------CACAGTGACACA 763
                 :::       :::|||          ||||||                     |||:::
Db        373 GlyHisGlnSerLysSerMetLeuLysAspAspLeuLysLeuSerSerSerGluAspSer 392

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              |||    |||::::::    |||                  |||          :::
Db        393 AspGlyGluGlnAspCysAspLys------------ThrMetProArgSerThrProGly 408

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
                    :::             |||       |||          :::    :::
Db        409 SerAsnSerGluProSerHisHisAsnSerGluGlyAlaAspAsnSerArgAspAspSer 428

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              |||    |||    |||    :::    :::|||
Db        429 SerSerHisSerGlySerGluSerSerSer------------------------------ 438

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
                                    ||||||:::                |||    :::|||:::
Db        439 ---------------------GlySerAspSerGluSerGluSerSerSerAspSer 451

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
              :::                                   |||    |||             |||
Db        452 Glu------------------------AlaAsnGluProSerGlnSerAlaSerProGlu 463

Qy       1064 CCCGAATCACCG------------ATGGATGTGGATAAT--------------------- 1090
              ||||||    |||                      :::||||||
Db        464 ProGluProProProThrAsnLysTrpGlnLeuAspAsnTrpLeuAsnLysValAsnPro 483

Qy       1090 ------------------------------------------------------------ 1090

Db        484 HisLysValSerProAlaSerSerValAspSerAsnIleProSerSerGlnGlyTyrLys 503

Qy       1091 ---------------------TCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAG 1129
                                    :::    ||||||    |||:::          :::|||
Db        504 LysGluGlyArgGluGlnGlyThrGlyAsnSerTyrThrAspThrSerGlyProLysGlu 523

Qy       1130 ACAAGT---------CCAGGTTTTGATGAACAA------GAAGATGGTAGTTCCTCCCAA 1174
              ||||||          ||||||    |||    :::       :::    ||||||    |||
Db        524 ThrSerSerAlaThrProGlyArgAspSerLysThrIleGlnLysGlySerGluSerGly 543

Qy       1175 ACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAA--- 1231
                       :::    |||       |||:::    |||:::       :::    |||
Db        544 ArgGlyArgGlnLysSerProAlaGlnSer---AspSerThrThrGlnArgArgThrVal 562

Qy       1232 ---CGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGA--- 1285
                 :::          |||       |||                 :::    ||||||
Db        563 GlyLysLysGlnProLysLysAlaGluLysAlaAlaAlaGluGluProArgGlyGlyLeu 582
```

```
Qy     1286 ------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCT 1339
                  ||||||     |||      :::    ||||||   :::   :::||||   :::
Db      583 LysIleGluSerGluThrProVal---AspLeuAlaSerSerMetProSerSerArgHis 601

Qy     1340 AGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAAATCACA 1399
                     :::      :::    |||    :::    ||||||:::||||||   :::
Db      602 LysAlaAlaThrLys---GlySerArgLysProAsnIleLysLysGluSerLysSerSer 620

Qy     1400 GATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAA 1459
                           |||   ||||||    ::::::|||    :::       :::|||
Db      621 ProArg------------ProThrAlaGlu---LysLysLysTyrLysSerThrSerLys 635

Qy     1460 CATCAAAGAACAGAAAGGAAGATC------------------------------------ 1483
                  :::      |||:::|||
Db      636 SerSerGlnLysSerArgGluIleIleGluThrAspThrSerSerSerAspSerAspGlu 655

Qy     1484 ------------------------------CCTAAATACGTTCCACCTCACCTTTCTCCA 1513
                                          |||||||||      :::    :::|||
Db      656 SerGluSerLeuProProSerSerGlnThrProLysTyrProGluSerAsnArgThrPro 675

Qy     1514 GATAAG------------------------------------AAGTGGCTTGGAACTCCC 1537
               |||                                    :::   :::   :::|||
Db      676 ValLysProSerSerValGluGluGluAspSerPhePheArgGlnArgMetPheSerPro 695

Qy     1538 ATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
            :::||||||         :::                   ||||||      |||
Db      696 MetGluGluLysGluLeuLeuSer------------------ProLeuSerGluProAsp 709

Qy     1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCT---AAA 1654
               :::   :::   :::::::::||||||                :::|||    |||
Db      710 AspArgTyrProLeuIleValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLys 729

Qy     1655 CCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCA 1714
            |||            ||||||:::               |||       |||
Db      730 Pro------------TyrLysGluThrGluProProLysGlyGluLysLysAsnValPro 745

Qy     1715 GAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGG 1774
            |||::::::           :::   :::      |||:::   :::
Db      746 GluLysHisThrArgGluAlaGlnLysGlnAlaSerGluLysValSer------------ 761

Qy     1775 GAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
                                    ||||||      |||           :::|||
Db      762 ------------------------AsnLysGlyLysArgLysHisLysAsnGluAsp--- 772

Qy     1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTC 1894
                           :::    |||||||||
Db      773 ------AspAsnArgAlaSerGluSerLysLys-------------------------- 781

Qy     1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTA------------TATCAGTCC 1942
                   |||        ::::::         |||   |||                :::|||
Db      782 ---ProLysThrGluAspLysAsnSerAlaGlyHisLysProSerSerAsnArgGluSer 800

Qy     1943 ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 2002
                           |||            ||||||:::               |||:::
Db      801 SerLysGlnSerAlaAlaLysGluLysAspLeuLeuProSer---ProAlaGlyProVal 819

Qy     2003 CCACTCCTGAAACAGAAGATGAATCATTCCATC---------ACAATGTCGCAGGAACAG 2053
            |||            |||      |||               |||:::||||||
Db      820 ProSerLysAspProLysThrGluHisGlySerArgLysArgThrIleSerGlnSerSer 839

Qy     2054 ATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATG 2113
                              :::|||:::                      |||   :::
Db      840 SerLeuLysSerSerSerAsnSer------------------------AsnLysGluThr 851

Qy     2114 AAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCA 2173
                              :::            :::          ||||||::::::
Db      852 SerGlySerSerLysAsnSerSerSerThrSerLysGlnLysLysThrGluGlyLysThr 871

Qy     2174 TCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAA 2233
            |||         :::::::::|||                              ||||||
Db      872 SerSerSerSerLysGluValLys------------------------------GluLys 881

Qy     2234 AAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAA 2293
              |||:::                                  |||    :::   |||
Db      882 AlaProSer------------------------SerSerSerAsnCysProPro------ 891

Qy     2294 AGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAAT 2353
                  |||                                    |||::::::   :::
Db      892 ------SerAlaPro------------------------------ThrLeuAspSerSer 899

Qy     2354 GGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGT 2413
                     ::::::   |||    :::|||                       ::::::
Db      900 LysProArgArgThrLysLeuValPheAspAspArg------------AsnTyrSerAla 915

Qy     2414 GCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGG 2473
                     :::||||||      :::         |||   :::   :::      |||
Db      916 AspHisTyrLeuGlnGluAlaLysLysLeuLysHisAsnAlaAspAlaLeuSerAspArg 935

Qy     2474 CTC-----------------TTCACTGAGGTGCTGGATCACAATGAATGT--------- 2506
                                         |||:::            ||||||
Db      936 PheGluLysAlaValTyrTyrLeuAspAlaValValSerPheIleGluCysGlyAsnAla 955

Qy     2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACA------ 2560
             |||              |||      |||      :::      |||:::||||||
Db      956 LeuGluLysAsnAlaGlnGluSerLysSerProPheProMetTyrSerGluThrValAsp 975

Qy     2561 ---------TATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAG 2611
                     |||               :::                  :::         :::
Db      976 LeuIleLysTyrThrMetLysLeuLysAsnTyrLeuAlaProAspAlaThrAlaAlaAsp 995

Qy     2612 CGGCGCTGCACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAG 2671
             :::|||   |||     :::     |||    ||||||     ||||||     |||
Db      996 LysArgLeuThrValLeu--CysLeuArgCysGluSerLeuLeuTyrLeuArgLeuPheL 1015

Qy     2672 TTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT--- 2728
                       ||||||         :::     :::|||               |||:::
Db     1015 ysLeuLysLysGluAsnAlaLeuLys-TyrSerLysThrLeuThrGluHisLeuLysAsn 1034

Qy     2729 --------------------------CCTGGAGTTTCTTCAGAG--------------- 2746
                                      ||||||:::   |||:::
Db     1035 SerTyrAsnAsnSerGlnAlaProSerProGlyLeuGlySerLysAlaValGlyMetPro 1054

Qy     2747 ---------------AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTT 2791
                              ||||||         :::||||||:::         ||||||
Db     1055 SerProValSerProLysLeuSerProGlyAsnSerGlyAsnTyrSerSerGlyAlaSer 1074

Qy     2792 GGGGGTGATGCC 2803
                     |||
Db     1075 SerAlaSerAla 1078
```

RESULT 13
US-09-815-242-12955
; Sequence 12955, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

```
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12955
;   LENGTH: 1111
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12955

Alignment Scores:
Pred. No.:               9.26e-05     Length:         1111
Score:                   175.00       Matches:        111
Percent Similarity:      34.62%       Conservative:   68
Best Local Similarity:   21.47%       Mismatches:     171
Query Match:             2.43%        Indels:         167
DB:                      9            Gaps:           19

US-09-302-812-3 (1-4069) x US-09-815-242-12955 (1-1111)

Qy        311 AGGGTCCCACCGTCCTCGCCAGCC------------------------------------ 334
              |||:::|||   |||   |||:::
Db        203 ArgIleProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222

Qy        335 ---TGCGTCCCAGGGCAGGCGGGACAGCACAGA------------GGCAGCGCCACCTCG 379
                           |||     |||::::::
Db        223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGluGlnGluLysGlnAsnAsp 242

Qy        380 CTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCG 439
                 |||   ||||||   |||    :::         |||                |||
Db        243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerSerThrThr 262

Qy        440 GAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGT 499
              :::|||             ||||||   :::|||    :::||||||::::::   :::
Db        263 LysSerMetProAsnTyrSerLysValAspAsnThrIleLysIleGluAsnIleTyrAla 282

Qy        500 TCT------------------------------------------GTACAAAAA------ 511
              |||                                          :::||||||
Db        283 SerGlnIleValGluGluIleArgArgGluArgGluArgLysValLeuGlnLysArgArg 302

Qy        511 ------------------------------------------------------------ 511

Db        303 PheLysLysAlaLeuGlnGlnLysArgGluGluHisLysAsnGluGluGlnAspAlaIle 322

Qy        512 ------------GATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAG 559
                          |||      |||          |||:::   |||      |||
Db        323 GlnArgAlaIleAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342

Qy        560 CTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACT 619
              |||:::   |||      ||||||:::   ||||||                  |||
Db        343 LeuAsnAspAspSerAspLeuThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362

Qy        620 GCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAA 679
                                                                      |||
Db        363 GlyIle---------------------------------------------------Glu 365

Qy        680 CCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCT 739
                 :::||||||:::                     ::::::   |||      |||
Db        366 AsnGluThrValSer------------------AsnAspGluAsnLysGlnAlaSerIle 379

Qy        740 CAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTC 799
              ||||||:::|||   :::||||||       :::|||:::        |||   :::   :::
Db        380 GlnAsnGluAspThrAsnAspThrHisLeuAspGluSerProTyrAsnTyrGluGluVal 399

Qy        800 ACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAA------CACGCCAGA 853
              :::      :::::::::      ||||||   ::::::||||||         :::
Db        400 SerLeuAsnGlnValSerThrThrLysGlnLeuSerAspAspGluValThrValSerAsn 419

Qy        854 GAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGT 913
                             ||||||                                    |||
Db        420 ValThrSerGlnHisGln------------------------------------Ser 426

Qy        914 TGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
                 |||         ::::::|||      |||      |||
Db        427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu------------------ 440

Qy        974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
                                 ||||||   ::::::||||||                  |||
Db        441 ------------------LysAsnGlnSerArgLeuIleAlaAspSerGluGluAspGly 454

Qy       1034 AATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCT 1093
                 :::             ||||||            |||   :::|||   |||
Db        455 AlaThr---------AsnLysGluGluTyrSerGlySerGlnIleAsp---AspAlaGlu 470

Qy       1094 AAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAA 1153
                             |||:::|||   ||||||:::|||:::          ::::::
Db        471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnIleGluAspAsn 490

Qy       1154 GAAGATGGTAGTTCCTCCCAA---------ACAGCAAATAAACCTTCAAGGTTC------ 1198
                 :::   ::::::|||:::               |||   |||      :::
Db        491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluTyrAlaVal 510

Qy       1199 ---------------------------CAAGCAAGAGACGCTGACATTGAATTTAGGAAA 1231
                                         :::      |||   :::||||||
Db        511 ThrGluSerGlnValAsnAsnIleAspLysThrValAspAsnGluIleGluLeuAlaPro 530

Qy       1232 CGGTAC---------------------------------TCTACTAAGGGCGGTGAAGTT 1258
              |||:::                                 |||   |||      :::|||
Db        531 ArgHisLysLysAspAspGlnThrAsnLeuSerValAsnSerLeuLysThrAsnAspVal 550

Qy       1259 AGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCT 1318
                 :::      |||                     :::   ||||||::::::
Db        551 AsnAspAsnHisValValGluAsp------------SerSerMetAsnGluIleGluLys 566

Qy       1319 AAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
               |||                            ||||||||||||      |||::::::
Db        567 -Asn----------------------------AsnAlaGluIleThrGluAsnValGl 576

Qy       1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
              :            :::|||:::                          |||      |||||
Db        576 nAsnGluAlaAlaGluSerGlu-----------------------GlnAsnValGluGl 588

Qy       1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA 1498
              |||| |||:::   |||            |||   :::   |||               |||
Db        588 uLys-ThrIleGluAsnValAsnProLysLysGlnThrGluLysValSerThrLeuSerL 608

Qy       1499 CCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATG 1558
                   ||||||:::         |||         ||||||   ::::::   |||   |||
Db        608 ysArgProPheAsn---------ValValMet-ThrProSerAspLysLysArgMetMet 624

Qy       1559 CCTCGG---------TGTGGGATCCGGCTGCCTCTCTTGAGACCA 1594
                 |||              :::   :::|||   |||:::|||
Db        625 AspArgLysLysHisSerLysValAsnValProGluLeuLysPro 639

RESULT 14
US-10-282-122A-43837
; Sequence 43837, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
```

```
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43837
;   LENGTH: 1274
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-10-282-122A-43837

Alignment Scores:
Pred. No.:               0.000118       Length:         1274
Score:                   174.00         Matches:        112
Percent Similarity:      35.28%         Conservative:   69
Best Local Similarity:   21.83%         Mismatches:     173
Query Match:             2.41%          Indels:         159
DB:                      12             Gaps:           20

US-09-302-812-3 (1-4069) x US-10-282-122A-43837 (1-1274)

Qy      311 AGGGTCCCACCGTCCTCGCCAGCC---------------------------------- 334
            |||:::|||   |||   |||:::
Db      203 ArgIleProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222

Qy      335 ---TGCGTCCCAGGGCAGGCGGGACAGCACAGA-----------GGCAGCGCCACCTCG 379
                              |||      |||::::::
Db      223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGluGlnGluLysGlnAsnAsp 242

Qy      380 CTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCG 439
               |||   ||||||  |||   :::             |||                |||
Db      243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerSerThrThr 262

Qy      440 GAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGT 499
            :::|||             ||||||   :::|||      :::||||||::::::   :::
Db      263 LysSerMetProAsnTyrSerLysValAspAsnThrIleLysIleGluAsnIleTyrAla 282

Qy      500 TCT---------------------------------------GTACAAAAA------ 511
            |||                                      :::||||||
Db      283 SerGlnIleValGluGluIleArgArgGluArgGluArgLysValLeuGlnLysArgArg 302

Qy      511 ------------------------------------------------------------ 511

Db      303 PheLysLysAlaLeuGlnGlnLysArgGluGluHisLysAsnGluGluGlnAspAlaIle 322

Qy      512 ------------GATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAG 559
                        |||   |||         |||:::   |||          |||
Db      323 GlnArgAlaIleAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342

Qy      560 CTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACT 619
            |||:::   |||         ||||||:::   ||||||              |||
Db      343 LeuAsnAspAspSerAspLeuThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362

Qy      620 GCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAA 679
                                                                     |||
Db      363 GlyIle---------------------------------------------------Glu 365

Qy      680 CCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCT 739
               :::||||||:::                  ::::::   |||      |||
Db      366 AsnGluThrValSer------------------AsnAspGluAsnLysGlnAlaSerIle 379

Qy      740 CAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTC 799
            ||||||:::|||   :::||||||   :::|||:::         |||   :::   :::
Db      380 GlnAsnGluAspThrAsnAspThrHisValAspGluSerProTyrAsnTyrGluGluVal 399

Qy      800 ACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAA------CACGCCAGA 853
            :::      :::::::::      ||||||   ::::::||||||         :::
Db      400 SerLeuAsnGlnValSerThrThrLysGlnLeuSerAspAspGluValThrValSerAsn 419

Qy      854 GAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGT 913
                     ||||||                                          |||
Db      420 ValThrSerGlnHisGln---------------------------------------Ser 426

Qy      914 TGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
                  |||      ::::::|||      |||      |||
Db      427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu------------------ 440

Qy      974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
                              ||||||   ::::::||||||                  |||
Db      441 ------------------LysAsnGlnSerArgLeuIleAlaAspSerGluGluAspGly 454

Qy     1034 AATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCT 1093
            :::                   ||||||         |||   :::|||   |||
Db      455 AlaThr---------AsnLysGluGluTyrSerGlySerGlnIleAsp---AspAlaGlu 470

Qy     1094 AAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAA 1153
                     |||:::|||   ||||||:::|||:::            ::::::
Db      471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnIleGluAspAsn 490

Qy     1154 GAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGA------ 1207
               :::   ::::::|||:::                  :::   |||
Db      491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluHisAlaVal 510

Qy     1208 ------------------------------------GACGCTGACATTGAATTTAGGAAA 1231
                                                |||   :::||||||
Db      511 ThrGluSerGlnValAsnAsnIleAspLysThrValAspAsnGluIleGluLeuAlaPro 530

Qy     1232 CGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGT 1291
            |||:::      |||         :::      |||:::                  :::
Db      531 ArgHisLys---LysAspAspGlnThrAsnLeuAsnVal------------------Asn 543

Qy     1292 CGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATT---------TCTAGC 1342
                     ||||||:::|||               |||::::::         ||||||
Db      544 SerLeuLysThrAsnAspValAsn---------AspGlyHisValValGluAspSerSer 560

Qy     1343 CTGAAT------------GTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCT 1390
            :::|||            ||||||||||||         |||::::::
Db      561 MetAsnGluIleGluLysHis-AsnAlaGluIleThrGluAsnValGlnAsnGluAlaAl 580

Qy     1391 AAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGG 1450
              :::|||:::                         |||      |||||||||  |||:::
Db      580 aGluSerGlu------------------------GlnAsnValGluGluLys-ThrIleG 592

Qy     1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCT 1510
             |||          |||   :::   |||                 |||      ||||||:
Db      592 luAsnValAsnProLysLysGlnThrGluLysValSerThrLeuSerLysArgProPheA 612

Qy     1511 CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGG------ 1564
            ::         |||      ||||||   ::::::      |||   |||   |||
Db      612 sn---------ValValMet-ThrProSerAspLysLysArgMetMetAspArgLysLys 628

Qy     1565 ---TGTGGGATCCGGCTGCCTCTCTTGAGACCA 1594
                   :::   :::|||   |||:::|||
Db      629 HisSerLysValAsnValProGluLeuLysPro 639
```

```
RESULT 15
US-09-864-761-43244
; Sequence 43244, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43244
;   LENGTH: 748
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004886.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;   OTHER INFORMATION: EST_HUMAN HIT: AI636014.1, EVALUE 3.00e-51
;   OTHER INFORMATION: SWISSPROT HIT: P38254, EVALUE 1.40e+00
US-09-864-761-43244

Alignment Scores:
Pred. No.:              0.000104      Length:         748
Score:                  173.50        Matches:        133
Percent Similarity:     37.54%        Conservative:   117
Best Local Similarity:  19.97%        Mismatches:     239
Query Match:            2.41%         Indels:         177
DB:                     9             Gaps:           30

US-09-302-812-3 (1-4069) x US-09-864-761-43244 (1-748)
```

```
Qy      320 CCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCG 379
             |||  ||||||         :::                   |||   :::::|||
Db       15 ProLysSerProGlnGlyLeuSerAspThrGlyTyrSerSerAspGlyIleSerSerSer 34

Qy      380 CTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT------------------ 421
             |||                ||| |||  :::  |||
Db       35 Leu------------GlyGluIleProSerLeuIleProThrAspGluLysAspIleLeu 50

Qy      422 AAAGGAATCAAG------ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAAT 475
             ||||||:::|||        :::          ||||||      |||      ::::::
Db       51 LysGlyLeuLysLysAspSerPheSerGlnGluSerSerProSerSerProSerAspLeu 70

Qy      476 ACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTA 535
               ::::::||||||   :::   |||:::   :::
Db       71 AlaLysLeuGluSerThrValLeuSerIleLeuGlu------------------------ 82

Qy      536 GAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACA 595
                                |||   |||:::   :::||||||        ||||||  |||
Db       83 ---------AlaGlnAlaSerThrLeuAlaAspGluLysSer------GluLysLysThr 97

Qy      596 CAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACAC 655
             |||                 |||:::   :::              :::      :::   :::
Db       98 Gln---------ProHisGluValSerPro------GluGlnProLysAspGlnGluLys 112

Qy      656 ACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGT 715
             |||:::   |||       ||||||        :::|||:::       |||:::
Db      113 ThrGlnSerLeu------SerGluThrLeuGluIleThrIleSerGluGluGluIleLys 130

Qy      716 AATGCTAACATTGATCGG------------------TCACCTCAAAATGATGATCACAGT 757
               :::      :::|||                     |||   |||     |||
Db      131 GluSerGlnGluGluArgLysAspThrPheLysLysAspSerGlnGlnAspIleProSer 150

Qy      758 GACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------ 811
                   |||   :::|||           :::        :::||||||   :::
Db      151 SerLysAspHisLysGluLysSerGluPheValAspAspIleThrThrArgArgGluPro 170

Qy      812 CTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCC------------AGAGAAGCC 859
               :::   ::::::::::::|||   |||::::::              |||:::
Db      171 TyrAspSerValGluGluSerSerGluSerGluAsnSerProValProGlnArgLysArg 190

Qy      860 AAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAG 919
             ::::::            |||   |||             ::::::   :::
Db      191 ArgThrSerValGlySerSerSerSerAspGluTyrLysGlnGluAspSerGlnGlySer 210

Qy      920 CAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTT 979
               :::|||   |||   :::   |||         :::    ::::::   :::::::::
Db      211 GlyGluGluGluAspPheIleArgLysGln---IleIleGluMetSerAlaAspGluAsp 229

Qy      980 GGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCT 1039
               :::||||||::::::|||               |||          :::   :::|||
Db      230 AlaSerGlySerGluAspAspGluPheIleArgAsnGlnLeuLysGluIleSerSerSer 249

Qy     1040 CCTCCATTTGAGAAGGAAAGTGAACCCGAA------------------------------ 1069
                  :::|||:::   |||   :::
Db      250 ThrGluSerGlnLysLysGluGluThrLysGlyLysGlyLysIleThrAlaGlyLysHis 269

Qy     1070 ------------------------TCACCGATGGATGTGGATAATTCTAAAAATAGTTGT 1105
                                     :::   :::|||   |||         :::
Db      270 ArgArgLeuThrArgLysSerSerThrSerIleAspGluAspAlaGlyArgArgHisSer 289

Qy     1106 CAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGT 1165
                      |||   ||||||               |||||||||       :::
Db      290 TrpHisAspGluAspAspGlu------------AlaPheAspGluSerProGluLeuLys 305

Qy     1166 TCCTCCCAAACAGCAAATAAACCTTCA--------------------------------- 1192
                :::|||   ::::::   |||
Db      306 TyrArgGluThrLysSerGlnGluSerGluGluLeuValValThrGlyGlyGlyGlyLeu 325

Qy     1193 ---AGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGC 1249
               ||||||:::        :::   :::                  :::||||||   :::
Db      326 ArgArgPheLysThrIleGluLeuAsnSerThrIleAlaAspLysTyrSerAlaGluSer 345
```

```
Qy     1250 GGTGAAGTTAGATTACATTTCCAATTT--------------------------------- 1276
              :::   :::            |||
Db      346 SerGlnLysLysThrSerLeuTyrPheAspGluGluProGluLeuGluMetGluSerLeu 365

Qy     1277 ---GAAGGAGGAGAGAGTCGCACT---------GGAATGAATGATTTAAATGCTAAA--- 1321
              :::       |||   |||:::         |||   :::   |||:::|||
Db      366 ThrAspSerProGluAspArgSerArgGlyGluGlySerSerSerLeuHisAlaSerSer 385

Qy     1322 ---------------CTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
                           |||   :::::|||||||||:::   :::      :::|||
Db      386 PheThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSerAspSerSer 405

Qy     1367 AAGCAACATGGAAAAAAGGATTCTAAA-----------------ATCACAGATCATTTG 1408
                |||   |||   :::||||||                           |||
Db      406 ProSerHisLysLysGlyGluSerLysGlnGlnArgLysAlaArgHisArgProHisGly 425

Qy     1409 ATGAGACTGCCCAAAGCAGAGGACAGA-----------------AGAAAAGAACAGTGG 1450
                ||||||      ||||||                         |||:::|||:::
Db      426 ProLeuLeuProThrIleGluAspSerSerGluGluGluLeuArgGluGluGluGlu 445

Qy     1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATC-------------------------- 1483
                |||   |||      :::|||:::|||
Db      446 LeuLeuLysGluGlnGluLysGlnArgGluIleGluGlnGlnGlnArgLysSerSerSer 465

Qy     1484 ------------------------------------------------------CCTAAA 1489
                                                                  ||||||
Db      466 LysLysSerLysLysAspLysAspGluLeuArgAlaGlnArgArgArgGluArgProLys 485

Qy     1490 TACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATG 1549
                |||   :::|||||||||   :::                 :::|||   ||||||:::
Db      486 ThrProProSerAsnLeuSerProIleGluAsp------AlaSerProThrGluGluLeu 503

Qy     1550 AGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT---GCCAATCAC 1606
            |||:::                           |||   |||   |||   :::   :::
Db      504 ArgGlnAlaAlaGluMetGlu---------GluLeuHisArgSerSerCysSerGluTyr 520

Qy     1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
            :::   :::|||      |||                     |||:::   |||
Db      521 SerProSerIleGluSerAsp---------------------ProGluGlyPheGluIle 533

Qy     1667 CATTATAAAGATTTGTGGGAT---AACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
               :::   :::   :::         |||      |||:::|||   :::      :::
Db      534 SerProGluLysIleIleGluValGlnLysValTyrLysLeuPro---ThrAlaValSer 552

Qy     1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
            ||||||            ||||||
Db      553 LeuTyrSerProThrAspGluGln------------------------------------ 560

Qy     1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
                  ::::::::   |||               :::   ||||||   |||
Db      561 ------SerIleMetGlnLysGluGlySerGlnLysAlaLeuLysSerAlaGluGluMet 578

Qy     1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTG------ATCGATTTCTGG 1897
            |||               |||         :::   |||            :::
Db      579 TyrGluGluMetMetHisLysThrHisLysTyrLysAlaPheProAlaAlaAsnGluArg 598

Qy     1898 GATAAGGTACTTGAAGAAGCAGAA---------------------GCTCAACATTTATAT 1936
            |||:::|||   |||:::                               ::::::|||
Db      599 AspGluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAspTyrIleTyr 618

Qy     1937 CAGTCCATCTTGCCTGAT 1954
            :::|||::::::   |||
Db      619 GluSerLeuValGluAsp 624
```

Search completed: May 26, 2004, 18:36:27
Job time : 276.187 secs

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 26, 2004, 16:16:17 ; Search time 59.0623 Seconds
(without alignments)
13253.917 Million cell updates/sec

Title: US-09-302-812-3
Perfect score: 7208
Sequence: 1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
```
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150054_5689/app_query.fasta_1.12437
-DB=PIR_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_298_@runat_26052004_150054_5689 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	512	7.1	997	2	B84726	probable poly(ADP-
2	463	6.4	726	2	T21138	hypothetical prote
3	340	4.7	364	2	A84726	probable poly(ADP-
4	181.5	2.5	1359	2	T34036	hypothetical prote
5	180	2.5	1298	2	I54367	X-linked nuclear p
6	174	2.4	1274	2	A89959	hypothetical prote
7	171.5	2.4	884	2	D96730	unknown protein F5
8	168	2.3	3147	2	T18674	hypothetical prote
9	167	2.3	4717	2	T41581	hypothetical coile
10	166.5	2.3	1641	2	I38614	helicase II - huma
11	161	2.2	1788	2	T29043	hypothetical prote
12	160	2.2	1164	2	T24806	hypothetical prote
13	160	2.2	1957	2	T38077	hypothetical coile
14	159.5	2.2	1183	2	S65236	probable membrane
15	159	2.2	406	2	S38170	SRP40 protein - ye
16	159	2.2	1031	2	T38411	probable GTPase ac
17	159	2.2	1403	1	A47328	natural killer cel
18	159	2.2	2429	1	SJHUA	spectrin alpha cha
19	158.5	2.2	6642	2	T29757	protein UNC-89 - C
20	158	2.2	967	2	T41672	probable b-zip tra
21	158	2.2	1974	2	T30010	hypothetical prote
22	156.5	2.2	1538	2	T29095	cardiac muscle fac
23	156	2.2	7160	2	T27935	hypothetical prote
24	155	2.2	1200	2	A46194	neurofilament prot
25	154.5	2.1	646	1	S15901	chromogranin B pre
26	154.5	2.1	3225	2	I52300	giantin - human
27	154.5	2.1	3259	1	A56539	giantin - human
28	154	2.1	1938	2	JC5421	smooth muscle myos
29	154	2.1	1972	2	JC5420	smooth muscle myos
30	154	2.1	2218	2	B84683	hypothetical prote
31	153	2.1	1071	2	E85343	hypothetical prote
32	152	2.1	644	2	S55395	neurofilament prot
33	152	2.1	1320	2	S57113	BUD4 protein - yea
34	151	2.1	1272	2	C90593	hypothetical prote
35	151	2.1	3187	2	JC5837	364K Golgi complex
36	150.5	2.1	1312	1	BWBYDL	RAD50 protein - ye
37	150.5	2.1	1972	1	A41604	myosin heavy chain
38	150	2.1	1061	2	G88428	protein C28A5.6 [i
39	150	2.1	1206	2	T34021	protein kinase SK2
40	150	2.1	2137	1	SJHUB	spectrin beta chai
41	149.5	2.1	1210	2	I39410	AF-4 protein, spli
42	149.5	2.1	1213	2	A58198	serine/proline-ric
43	149	2.1	1093	2	F38556	protein B0464.5a [
44	149	2.1	5170	2	T15348	hypothetical prote
45	148.5	2.1	472	2	T27903	hypothetical prote

ALIGNMENTS

RESULT 1
B84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31870
A;Map position: 2

```
Alignment Scores:
Pred. No.:              1.47e-28      Length:         997
Score:                  512.00        Matches:        176
Percent Similarity:     41.20%        Conservative:   79
Best Local Similarity:  28.43%        Mismatches:     182
Query Match:            7.10%         Indels:         182
DB:                     2             Gaps:           19

US-09-302-812-3 (1-4069) x B84726 (1-997)

Qy       1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                 :::|||:::::::::::|||    |||:::::           :::       |||:::|||
Db         79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98

Qy       1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATT--------------------------- 1987
                 :::      :::    |||       |||::::::
Db         99 LeuAlaSerLeuLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118
```

```
Qy   1988 ---------------TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCC 2032
Db    119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138

Qy   2033 ATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTT 2092
Db    139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158

Qy   2093 CCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAAT 2152
Db    159 ProAspAspAsnArgGlyAlaLys-----------HisLeuProValIleAsnPheAsp 174

Qy   2153 CGA------TTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
Db    175 HisLeuSerLeuTyrIleSerTyrSerGlnSerGlnGluSerLysIleArgCysIleMet 194

Qy   2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
Db    195 HisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGluArgLys 214

Qy   2267 ------AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTG 2320
Db    215 IleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeu--------- 231

Qy   2321 CATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTT 2380
Db    232 ---------TyrGln-----------------ProAspAsnAlaLeuGluValAspPhe 242

Qy   2381 GCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGC 2440
Db    243 AlaAsnLysTyrLeuGlyGlyGlySerLeuSerArgGlyCysValGlnGluGluIleArg 262

Qy   2441 TTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAAT 2500
Db    263 PheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAspAspAsn 282

Qy   2501 GAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACA 2560
Db    283 GluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAlaSerSer 302

Qy   2561 TATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGC 2620
Db    303 PheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspProPheLysArgArgArg 322

Qy   2621 ACTGAGATCGTTGCCATCGATGCTCTT-----------------CACTTCAGACGCTAC 2662
Db    323 ThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLysAspIle 342

Qy   2663 CTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTT 2722
Db    343 Cys--------------------LeuLeuArgGluIleAsnLysAlaLeuCysGlyPhe 355

Qy   2723 CTCCGT------------------------------------------------------ 2728
Db    356 LeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAspAsnGlu 375

Qy   2728 ------------------------------------------------------------ 2728
Db    376 IleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThrAlaSer 395

Qy   2728 ------------------------------------------------------------ 2728
Db    396 HisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsnLeuIle 415

Qy   2729 ---------------CCTGGAGTTTCTTCAGAGAATCTT------TCTGCAGTGGCCACA 2767
Db    416 ArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyValAlaThr 435

Qy   2768 GGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATA 2827
Db    436 GlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIleGlnTrp 455

Qy   2828 TTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTT---TATTTCACCTTTGGGGACTCAGAA 2884
Db    456 LeuAlaAlaSerGlnThrArgArgProPheIleSerTyrTyrThrPhe-GlyValGlu-- 474

Qy   2885 TTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACTGTTGGA 2944
Db    474 ------------------------------------------------------------ 474

Qy   2945 GATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGCAGAAACTGTTCCACCC-- 3002
Db    475 ---------------------------AlaLeuArgAsnLeuAspGlnLeuProProTh 485

Qy   3003 ----------------CTGGACCAGACATCAAGC---------------TTTATCCATTC 3031
Db    485 rLysLeuSerSerArgLeuAspSerAlaAsnSerProSerLeuThrCysLeuMetHisSe 505

Qy   3032 ATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
Db    505 rHisThr---------------ValSerLeuHisPhe-ProThrTrpAlaIleArgMetG 520

Qy   3092 ACCTGAGGAGCCGAGCGAATAGCATCTCCTCCCACCTCCCA------CCAGAGACGTCCT 3145
Db    520 luLeuArgAlaAspLeuArgSerIleLeuGlnTyrLeuProLeuValAlaGlnSerSerS 540

Qy   3146 GTTTGAGCTGTCAGGTGTAATATATGAATTGACTTAAGTTAATATAAATGTGTACATAAT 3205
Db    540 erLeuValTrpProProSerValGluGluGluLeuGlnThrIleSerArgGlyProSerG 560

Qy   3206 CCACATTTGTAGTCAAGGACGCAATCTCTTCCACACATGTG---------------CAGT 3250
Db    560 luSerMetValAsnSerGlyGluAlaLeuAlaLeuHisIleThrAsnMetArgLysSerL 580

Qy   3251 TGTCAGTTGGTACATCTAAACTCCCTCCATCCTGACTCACGTGGACTTAGATATGTTTTG 3310
Db    580 euSerLeuAsnAlaSerAspLeuAlaProTyrAlaLeu-GlnGly------TyrGlyLeu 597

Qy   3311 TTTCTATTTTCTTCTATTTCA-----------------------------GTTTTTCAT 3340
Db    598 PhePheAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValValPro 617

Qy   3341 TCTTTGATGTTTATTTCTTTTGTCCATCAGATCTCTTGTGAAATCCCAT 3389
Db    618 AlaLeu-CysArgLeuLeuLeuGlnLeuProSerMetLeuGluLysHis 633
```

RESULT 2
T21138
hypothetical protein F20C5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T21138
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19381
A;Accession: T21138
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-726 <WIL>
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1
A;Experimental source: clone F20C5
C;Genetics:
A;Gene: CESP:F20C5.1
A;Map position: 4
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

Alignment Scores:

Pred. No.:	4.5e-25	Length:	726
Score:	463.00	Matches:	207
Percent Similarity:	38.11%	Conservative:	123
Best Local Similarity:	23.90%	Mismatches:	278
Query Match:	6.42%	Indels:	258
DB:	2	Gaps:	37

```
US-09-302-812-3 (1-4069) x T21138 (1-726)

Qy    599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
Db     27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy    659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
Db     47 GluSerLysGluAsp--------------------------ProGluGluProLysSer 57

Qy    719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
Db     58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy    779 --------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
Db     78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy    815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
Db     98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy    863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
Db    116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy    923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
Db    126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy    983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
Db    137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy   1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
Db    151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy   1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
Db    164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy   1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
Db    184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy   1196 TTCCAAGCAAGAGACGCTGACATTGAATTT--------------------AGGAAACGG 1234
Db    204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy   1235 TACTCTACTAAGGGC--------------------------------------------- 1249
Db    224 ValArgThrSerSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy   1250 GGTGAAGTTAGATTACAT-----------------------TTCCAATTTGAAGGAGGA 1285
Db    244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy   1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
Db    264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy   1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
Db    283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy   1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
Db    303 ProLys-----------------------IleValLeuPro------------GlnArg 310

Qy   1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
Db    311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy   1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
Db    331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy   1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
Db    351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy   1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657
Db    368 ------------------------------------------------------------ 368

Qy   1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
Db    369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy   1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
Db    375 ------------------------------------------------------SerGlu 376

Qy   1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
Db    377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy   1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTC 1894
Db    394 ------------------------------------------------------------ 394

Qy   1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
Db    395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy   1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
Db    412 Val--------------------------------------------------------- 412

Qy   2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
Db    413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy   2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
Db    429 MetPhePhe---------------------AlaArgProAspSerProPheSer------ 439

Qy   2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
Db    440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy   2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
Db    456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy   2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
Db    476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy   2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
Db    494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy   2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
Db    514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy   2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
Db    534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy   2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
Db    554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573
```

```
Qy       2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
              |||||||||||        |||    :::|||                :::   |||   :::
Db        574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy       2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
              :::   :::|||   ||||||          |||         |||   :::|||||||||
Db        594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy       2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
              |||:::   |||:::                              :::|||
Db        614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              :::   ||||||:::   ||||||         ||||||:::         |||
Db        634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              |||:::      :::   ||||||   ||||||||||||||||||   ||||||
Db        654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||   :::||||||::::::||||||         |||:::|||   :::   :::
Db        673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::||||||:::   ||||||                  :::         :::   :::
Db        693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy       2927 AGGAAACTCACTGTTGGA 2944
              :::   :::|||:::|||
Db        713 LysAspValThrLeuGly 718
```

RESULT 3
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

```
Alignment Scores:
Pred. No.:              2.6e-16        Length:         364
Score:                  340.00         Matches:        104
Percent Similarity:     42.08%         Conservative:   58
Best Local Similarity:  27.01%         Mismatches:     119
Query Match:            4.72%          Indels:         104
DB:                     2              Gaps:           11

US-09-302-812-3 (1-4069) x A84726 (1-364)

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              :::   :::      |||:::                  ||||||   |||         :::
Db          1 MetSerLysGluGluSerSerArgTrpPheAsnGluPheLeuProAlaMetAlaCysLeu 20

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAG------------------------------ 1996
                 |||      |||::::::
Db         21 LeuLeuArgPheProSerLeuLeuGluSerHisTyrLeuAsnSerAspAsnLeuIleAsn 40

Qy       1997 ------------CCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCG 2044
                          :::   :::|||                     :::   :::|||
Db         41 GlyThrLysThrGlyLeuArgValLeuValProAsnLysAlaGlyIleValPheLeuSer 60

Qy       2045 CAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAAT 2104
              ||||||   |||   :::|||||||:::   :::|||||||||   ||||||
Db         61 GlnGluLeuIleGlyAlaLeuLeuSerCysSerPhePheCysLeuPhePro--------- 77

Qy       2105 GCTAAGATGAAATCGGAGTATTCTAGTTAC---CCAGACATTAACTTCAATCGATTGTTT 2161
                    :::         |||::::::   |||   |||||||||::::::||||||
Db         78 ------ValAspAspArgGlySerAsnHisLeuProIleIleAsnPheAspLysLeuPhe 95

Qy       2162 ---------------GAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
                             ||||||:::         :::         |||:::|||   :::
Db         96 GlySerLeuIleAsnThrGlyArgAsnGluHisGlnGluAsnLysIleLysCysIleIle 115

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
                 ||||||:::|||::::::                  |||   |||:::|||   |||:::
Db        116 HisTyrPheGlnArgLeuSerSerSerIleSerProGlyPheValSerPheGluArgLys 135

Qy       2267 ------AGTCTTGAA-----------------GATTTTCCAGAATGGGAAAGATGTGAA 2302
                    |||||||||                   |||         |||   :::
Db        136 IleLeuSerLeuGluGlnAspSerSerThrLeuAspGluGlyPheTrpGlyLysSerThr 155

Qy       2303 AAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGC 2362
                 |||      :::   |||                  |||   ||||||:::
Db        156 ValAsnLeuCysProValGluValArgThrSerGlyLeuIleGluAspGlnSerValGlu 175

Qy       2363 ATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTT 2422
                 |||:::||||||||||||||||:::   :::|||||||||
Db        176 AlaLeuGluValAspPheAlaAsnLysAsnLeuGlyGlyGlyAlaLeuArgLysGlyCys 195

Qy       2423 GTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACT 2482
              ||||||||||||||||||||||:::||||||||||||||||||:::         ||||||
Db        196 ValGlnGluGluIleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeu 215

Qy       2483 GAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATAC 2542
                 ::::::   |||   :::   :::   |||   |||::::::|||   |||
Db        216 ProThrMetGluValThrGluAlaIleGluValValGlyAlaGluArgPheSerLeuTyr 235

Qy       2543 ACAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGAC 2602
              ||||||
Db        236 ThrGly------------------------------------------------------ 237

Qy       2603 GACTGCGAGCGGCGCTGC---ACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGC 2659
                 |||   |||:::      |||   ||||||||||||||||||||
Db        238 ---CysPheArgLysAlaLysThrArgIleValAlaIleAspAlaLeuArg--------- 253

Qy       2660 TACCTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGA 2719

Db        253 ------------------------------------------------------------ 253

Qy       2720 TTTCTCCGTCCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGC 2779
                    ||||||||||||
Db        254 ------HisProGlyValSer--------------------------------------- 258

Qy       2780 TGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCT 2839
                                 :::|||::::::|||:::   :::|||         :::
Db        259 -----------------GlnTyrLysLeuGluSerLeuLeuSerValLeuIleLeuSer 272

Qy       2840 GCAGCTGAGCGAGATGTGGTTTATTTCACC---------------------TTTGGGGAC 2878
              ::::::   |||   :::      :::                           ||||||
Db        273 SerSerGlyArgProIleArgLeuTyrMetGlySerValSerLeuGlnGlyIleGlyAsp 292

Qy       2879 TCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACT 2938
                 ||||||   :::|||   |||      :::|||                 :::
Db        293 ValValLeuMetValGluIleLeuSerSerSerLeuPhePheAsnGlyLeuArgPheHis 312

Qy       2939 GTTGGAGATGTGTAT 2953
                 ::::::|||
Db        313 ArgSerAsnLeuTyr 317
```

RESULT 4
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans

```
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Alignment Scores:
Pred. No.:              8.45e-05     Length:          1359
Score:                  181.50       Matches:         182
Percent Similarity:     32.44%       Conservative:    131
Best Local Similarity:  18.86%       Mismatches:      328
Query Match:            2.52%        Indels:          324
DB:                     2            Gaps:            39

US-09-302-812-3 (1-4069) x T34036 (1-1359)

Qy     78 GAGAAGGACGCGCAGTCCATCTCTCTCAGGTTAGTGAAATGAGGCTCTCCGCGGGGCCGG 137
          ||||||   ||||||   :::      :::       :::    |||    |||
Db     33 GluLysArgAlaGlnLysLeuLysGluLysArgGluArgGluGlyLysPro--------- 49

Qy    138 CCCGGGGACAGTGCGCTGCTGGTCCCAGCATGAATGCGGGCCCCGGCTGTGAACCCTGCA 197
                                                                 |||
Db     50 ------------------------------------------------------ProPro 51

Qy    198 CCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCT 257
             ||||||||||||
Db     52 LysLysArgProAla--------------------------------------------- 56

Qy    258 TTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCC 317

Db     56 ------------------------------------------------------------ 56

Qy    318 CACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCT 377
             :::   |||:::   :::|||            |||     |||        |||
Db     57 ---LysLysArgLysAlaSerSerSerGluGluAspAspAspAspGluGluGluSerPro 75

Qy    378 CGCTTGTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAG 437
          |||     |||   |||                    :::  |||   ::::::  :
Db     76 ArgLysSerSerLysLys---------------------SerArgLysArgAlaLys--S 88

Qy    438 CGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGA 497
          ::||||||||||||   :::|||   |||                |||
Db     88 erGluSerGluSerAspGluSerAspGluGlu---------------GluAspArgLysL 103

Qy    498 GTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAA------AAATTAGTAAATG 551
             |||      |||           |||        |||     |||
Db    103 ysSerLysSerLysLysLysValAspGlnLysLysLysGluLysSerLysLysLysArgT 123

Qy    552 TTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGC 611
            :::       |||   |||:::          |||:::      |||
Db    123 hrThrSerSerSerGluAspGluAspSerAspGluGluArgGluGln------LysSerL 141

Qy    612 ATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGG 671
            :::   :::         |||   |||         :::    :::|||        |
Db    141 ysLysLysSerLysLysThrLysLysGlnThrSerSerGluSerSerGlu---------G 158

Qy    672 AAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATC 731
          ||||||||   :::                      ::::::         ::::::  :
Db    158 luSerGluGluGluArgLysValLysLysSerLysLysAsnLysGluLysSerValLysL 178

Qy    732 GGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAAC 791
             ::      :::      :::   ||||||   |||   :::   ::::::   :::::::
Db    178 ysArgAlaGluThrSerGluGluSerAspGluAspGluLysProSerLysLysSerLysL 198

Qy    792 AGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCCA 851
          ::    |||            |||   |||   ::::::      :::|||   |||    :
Db    198 ysGlyLeuLys------LysLysAlaLysSerGluSerGluSerGluSerGluAspGluL 216

Qy    852 GAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAA 911
          ::|||   |||      :::|||   ||||||                  |||      :::
Db    216 ysGluValLysLysSerLysLysLysSerLysLysValValLysLysGluSerGluSerG 236

Qy    912 GTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTG 971
                :::   |||                   |||        |||        ||||
Db    236 luAspGluAlaProGluLysLysLysThrGluLysArgLysArgSerLysThrSerSerG 256

Qy    972 AGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAG 1031
          ||:::                   |||:::|||   :::            :::|||
Db    256 luGluSerSerGluSerGluLysSerAspGluGlu---------GluGlu---------- 269

Qy   1032 GAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC------------GAATCACCGATGG 1079
                           ||||||||||||   |||                :::  |||:::
Db    270 ----------------GluLysGluSerSerProLysProLysLysLysLysProLeuA 284

Qy   1080 ATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCA- 1138
             |||            :::       :::       |||:::|||   |||      |||
Db    284 laValLysLysLeuSerSerAspGluGluSerGluGluSerAspValGluValLeuProG 304

Qy   1139 ------------------------------------GGTTTTGATGAACAAGAAGATG 1160
                                                  ||||||::::::
Db    304 lnLysLysLysArgGlyAlaValThrLeuIleSerAspSerGluAspGluLysAspGlnL 324

Qy   1161 GTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTG 1220
             |||   |||:::   ::::::                :::          :::   |||
Db    324 ysSerGluSerGluAlaSerAspValGluGluLysValSerLysLysLysAla------- 341

Qy   1221 AATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAG 1280
                :::|||:::   |||::::::                                     |
Db    342 -----LysLysGlnGluSerSerGlu----------------------------SerG 350

Qy   1281 GAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCTA 1340
          ||    :::|||         |||         :::|||   |||                  |
Db    350 lySerAspSerSerGluGlySerIleThrValAsnArgLys-----------------S 364

Qy   1341 GCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA------AAAAAGGATTCTAAAA 1394
          ||         |||         |||:::   |||                  ||||||:
Db    364 erLysLysLysGluLysProGluLysLysLysLysGlyIleIleMetAspSerSerLysL 384

Qy   1395 TCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAA 1454
          ::                  :::          :::||||||      |||:::::::::      |||
Db    384 euGlnLysGluThrIleAspAlaGluArgAlaGluLysGluArgArgLysArgLeuGluL 404

Qy   1455 CCAAACATCAAAGA---------------------------------------------- 1468
            |||   :::
Db    404 ysLysGlnLysGluPheAsnGlyIleValLeuGluGluGlyGluAspLeuThrGluMetL 424

Qy   1469 -------------ACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCTCCAG 1514
                       ::::::||||||:::                |||         |||   ||||
Db    424 euThrGlyThrSerSerGlnArgLysLeuLysSerValVal---------LeuAspProA 441

Qy   1515 ATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCC 1574
          ||                 :::   ::::::|||         :::   |||
Db    441 sp------------SerSerThrValAspGluGluSerLysLysPro------------- 452

Qy   1575 GGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGC 1634
                                                  |||    :::          |||:::|
Db    453 -----------------------------------ValGluValHisAsnSerLeuValA 461

Qy   1635 GAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGC 1694
          ||             :::   ||||||         |||
Db    461 rg---------IleLeuLysProHisGlnAlaHis------------------------- 469

Qy   1695 ATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGC 1754
                           |||     :::|||                   |||:::      :::|
```

```
Db        470 -----------------GlyIleGlnPheMetTyrAspCysAlaPheGluSerLeuAspA 484

Qy       1755 GA------ACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTA 1808
              ||       |||    ||||||
Db        484 rgLeuAspThrGluGlySer--------------------------------------- 490

Qy       1809 CACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAAT 1868
                                       ||||||   :::    :::          |||
Db        491 --------------------GlyGlyIleLeuAlaHisCysMetGlyLeuGlyLysThr- 503

Qy       1869 GGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAAGCTCAAC 1928
                       :::|||   |||         ||||||
Db        504 --------LeuGlnValIleThrPheLeuHisThrValLeu------------------- 514

Qy       1929 ATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCA---AATA 1985
                :::::::::   |||              ::::::   :::   :::|||   |||:
Db        515 --MetHisGluLysIleGlyGluLysCysLysArgValLeuValValValProLysAsnV 534

Qy       1986 TTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGC 2045
              ::                            ::::::          :::      :::
Db        534 alIleIleAsnTrpPhe------------LysGluPheGlnLysTrpLeuValAspAsnA 550

Qy       2046 AGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATG 2105
                |||::::::   ::::::       |||
Db        550 spGluGluLeuAspThrIleAspValAsn------------------------------ 559

Qy       2106 CTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGA------TTGT 2159
                                |||     ||||||      |||      |||      |||
Db        560 --------------GluLeuAspSerTyrLysThrIleGluAspArgArgArgAlaLeuL 575

Qy       2160 TTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAA 2219
                           ||||||:::   |||       :::         :::      ||||||
Db        575 ysAlaTrpHisSerSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgI 595

Qy       2220 GAGTCACA------------------GAGAAAAAACCTACTGGGTTGGTGACATTTACAA 2261
                :::|||                  :::|||||||||
Db        595 leLeuThrValGluAspAspProLysLysLysLysProLysAsnArgAsnArgArgLeuG 615

Qy       2262 GACAGAGTCTTGAAGATTTT-----------------------CCAGAATGGGAAAGAT 2297
                ::::::   |||||||||                      |||:::          |
Db        615 luLysAlaLysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValC 635

Qy       2298 GTGAA------------------------------------------------------AAAC 2306
              ||:::                                                   |||
Db        635 ysAspGluAlaHisLysLeuLysAsnAspAspSerAlaLeuSerLysCysMetValLysI 655

Qy       2307 CCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGC---------- 2356
                ||||||:::      :::         ||||||      :::|||
Db        655 leLeuThrLysArgArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluT 675

Qy       2357 -------------------------CAAGGCATGCTA------CAGGTGGATTTTGCAA 2384
                                        |||:::|||      :::   :::|||||||
Db        675 yrHisCysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaA 695

Qy       2385 ATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAA------------G 2432
              ||||||||||          :::   :::   |||       ::::::            |
Db        695 snArgPheVal------AsnIleIleAsnArgGlyArgThrLysAspAlaSerProLeuG 713

Qy       2433 AAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGG 2492
              ||:::   |||:::                  ::::::         :::   :::|
Db        713 luValSerPheMetLysArgArgCysHisValLeuTyrAspHisLeuLysLysCysValA 733

Qy       2493 ATCACAATGAATGTCTAATTATCACAGGTACT--------------------------- 2524
              ||       :::      ::::::|||
Db        733 spArgLysAspTyrArgValLeuThrGluAlaIleProProLysGlnGluTyrValIleA 753

Qy       2525 -----------------GAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTT 2567
                                |||   :::   |||      |||   |||      ||||||
Db        753 snValArgGlnThrGluArgGlnCysAlaLeuTyr-AsnAlaPheLeuAsnAspIleVal 772

Qy       2568 GGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGA 2627
              |||                                    ||||||   :::   |||
Db        773 GlyAsp------------------------------------SerGlyLeuSerLysArg 780

Qy       2628 TCGTTGCCA 2636
                 ||||||
Db        781 LeuLeuPro 783
```

RESULT 5
I54367
X-linked nuclear protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I54367
R;Gecz, J.; Pollard, H.; Consalez, G.; Villard, L.; Stayton, C.; Millasseau, P.; Khrestc
Hum. Mol. Genet. 3, 39-44, 1994
A;Title: Cloning and expression of the murine homologue of a putative human X-linked nuc
A;Reference number: I54367; MUID:94214473; PMID:8162050
A;Accession: I54367
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1298 <RES>
A;Cross-references: GB:L34363; NID:g508653; PIDN:AAA20872.1; PID:g530788
C;Genetics:
A;Gene: XNP

```
Alignment Scores:
Pred. No.:              0.000107      Length:          1298
Score:                  180.00        Matches:         199
Percent Similarity:     34.89%        Conservative:    160
Best Local Similarity:  19.34%        Mismatches:      374
Query Match:            2.50%         Indels:          296
DB:                     2             Gaps:            53
```

US-09-302-812-3 (1-4069) x I54367 (1-1298)

```
Qy        389 AAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448
              :::::::::   ::::::|||      :::||||||   |||::::::      |||
Db         17 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSerSer 36

Qy        449 AGTTTGGATAGTAAAGAAAAC------AACAATACAAGAATAGAATCCATGATGAGTTCT 502
              |||   |||::::::|||:::      :::|||   :::   :::          ||||||
Db         37 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysLysGlnArgThrSerSer 55

Qy        503 GTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA 562
                 :::|||                        :::         |||
Db         56 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 75

Qy        563 AGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCA 622
                          :::|||      |||:::      :::
Db         76 ArgLysGlnAlaAspIleThrSerSerSerSerSerAspIleGluAspAspAspGlnAsn 95

Qy        623 GCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCT 682
              ::::::                     ||||||      ||||||   :::      :::|||
Db         96 SerIle------------GlyGluGlySerSerAspGluGlnLysIle------LysPro 109

Qy        683 CAAACAGTAACCCTGGTACCAGAGCAG---------TTTAGTAATGCTAACATTGATCGG 733
                 |||      ||||||               |||      ::::::   |||
Db        110 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 129

Qy        734 TCA---------------------CCTCAAAATGATGATCACAGTGACACAGATAGTGAA 772
              :::                        :::||||||      |||   |||   |||
Db        130 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAsnAspProGlu 149

Qy        773 GAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------CTTGCAAATGCAAAG 826
                    :::      |||      :::|||      |||::::::   :::
Db        150 AsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuSerSerAspGlu 169

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
                 :::   |||:::         |||   |||
Db        170 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 189

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
                 :::||||||::::::   |||         |||   :::
Db        190 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer------------------ 203
```

```
Qy      947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
Db      204 ------GluSerAspSerAspSerGluAspLeu------ArgAlaLysIleGlnHisArg 219
Qy     1007 TTGATTAGACAAGAAAGTTGCCTA------GGAAATTCTCCTCCATTTGAGAAGGAAAGT 1060
Db      220 LeuPheGlyThrAsnTrpLeuLeuSerPheGlyGluSerGlyGlu---GluLysLysThr 238
Qy     1061 GAACCCGAATCACCGATGGATGTG---------GATAATTCTAAAAATAGTTGTCAAGAC 1111
Db      239 LysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSerGluAsp 258
Qy     1112 TCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCC 1171
Db      259 SerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal------SerGluSer 276
Qy     1172 CAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAA 1231
Db      277 GluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGluAsnGln 296
Qy     1232 CGG---TACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGA--- 1285
Db      297 ArgIleTyrLysGlnLysLysLysArgArgArgIleLysValGlnGluAspSerSerSer 316
Qy     1285 ------------------------------------------------------------ 1285
Db      317 GluAsnLysSerAsnSerGluGluGluGluGluGluLysGluGluGluGluGluGluGlu 336
Qy     1286 ------------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAAT 1333
Db      337 GluGluGluGluGluGluGluGluAspGluAsnAsp---AspSerLysSerProGly--- 354
Qy     1334 ATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAA------AAGGAT 1387
Db      355 -----------------------LysGlyArgLysLysIleArgLysIleLeuLysAsp 366
Qy     1388 TCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAAGAACAG 1447
Db      367 AspLysLeuArgThrGluThrGlnAsnAlaLeuLysGluGluGluArgLysGluArgArg 386
Qy     1448 TGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT 1507
Db      387 LysArgValArgGluArgGluThrGluLysLeuArgGluValIleGluIleGluAspAla 406
Qy     1508 TCTCCAGATAAGAAGTGGCTTGGAACTCCCATT---------------GAGGAGATGAGA 1552
Db      407 SerProThrLysCysProIleThrThrLysLeuValLeuAspGluAspGluGluThrLys 426
Qy     1553 AGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTA 1612
Db      427 Glu-----------------------ProLeuValGlnValHisArgAsnMetValIle 438
Qy     1613 ACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTAT 1672
Db      439 LysLeu--------------------------------LysPro---------HisGln 444
Qy     1673 AAAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAA 1720
Db      445 ValAspGlyValGlnPheMetTrpAsp---------------------CysCysCysGlu 457
Qy     1721 AATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTC 1780
Db      458 SerVal---------------LysLysThrLysLysSerProGlySerGlyCysIleLeu 472
Qy     1781 ATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTG 1840
Db      473 AlaHisCysMetGlyLeuGlyLys---------------ThrLeuGlnValValSerPhe 487
Qy     1841 AAATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTGATCGATTTCTGG 1897
Db      488 LeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeuValGlyLeuSer 507
Qy     1898 GATAAGGTACTT---------------GAAGAAGCAGAAGCTCAACATTTATATCAGTCC 1942
Db      508 SerSerIleLeuAlaPheAsnTrpMetAsnGluPheGluLys---------TrpGlnGlu 524
Qy     1943 ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 2002
Db      525 GlyLeuLysAspAspGluLysLeuGlu------ValSerGluLeuAlaThrValLysArg 542
Qy     2003 CCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGT 2062
Db      543 Pro---------GlnGluArgSerTyrMetLeuGlnArgTrpGlnGluAspGlyGlyVal 559
Qy     2063 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAG 2122
Db      560 MetIleIleGly------------------------------------------------ 563
Qy     2123 TATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAA 2182
Db      564 TyrGluMetTyr--------------ArgAsnLeuAlaGlnGlyArgAsnVal---Lys 577
Qy     2183 CCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACT 2242
Db      578 SerArgLysLeuLysGluIlePhe------AsnLysAlaLeuValAspProGlyProAsp 595
Qy     2243 GGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAA 2302
Db      596 PheValValCysAspGluGlyHisIleLeuLysAsn------GluAlaSerAlaValSer 613
Qy     2303 AAACCCTTG-----------ACACGATTGCATGTCACTTACGAAGGTACC--------- 2341
Db      614 LysAlaMetAsnSerIleArgSerArgArgArgIleIleLeuThrGlyThrProLeuGln 633
Qy     2342 -----------------------------------ATAGAAGAAAAATGGCCAAGGCATG 2365
Db      634 AsnAsnLeuIleGluTyrHisCysMetValAsnPheIleLysGluAsnLeuLeuGlySer 653
Qy     2366 CTACAGGTGGATTTTGCAAATCGTTTTGTT---------------GGAGGTGGTGTAACC 2410
Db      654 IleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGlyGlnCysAlaAsp 672
Qy     2411 AGT--------------------------------------------------------- 2413
Db      673 SerThrMetValAspValArgValMetLysLysArgAlaHisIleLeuTyrGluMetLeu 692
Qy     2414 GCAGGACTTGTGCAAGAAGAA---------------ATCCGCTTTTTA---ATCAATCCT 2455
Db      693 AlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLysPheLeuProProLysHis 712
Qy     2456 GAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTAATTATC 2515
Db      713 GluTyrValLeuAlaValArgMetThrSerIle------------GlnCysLysLeuTyr 728
Qy     2516 ACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAG------ACATATCGTTGG 2569
Db      729 AlaGlyAlaLysLeuPheGlnAspPheGlnMetLeuSerArgIleTrpThrHisProTrp 748
Qy     2570 --------------------TCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGC 2608
Db      749 CysLeuGlnLeuAspTyrIleSerLysGluAsnLysGlyTyrPheAspGluAspSerMet 768
Qy     2609 GAGCGGCGCTGCACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTC--- 2665
Db      769 Asp------------GluPheIleAlaSerAspSerAspGluThrSerMetSerLeuSer 784
Qy     2666 ---GATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTT 2722
Db      785 SerAspAspTyrThrLysLysLysLysLysGlyLysLysGlyLys--------------- 799
Qy     2723 CTCCGTCCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAAC--------- 2773
Db      800 --------------------LysAspSerSerSerSerGlySerGlySerAspAsnAsp 812
Qy     2774 ------------------TGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCC 2815
```

```
Db        813 ValGluValIleLysValTrpAsnSerArgSerArgGlyGlyGlyGlu------------ 828

Qy       2816 TTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTCACCTTTGGG 2875
                                                                      |||
Db        829 ---------------------------------------------------------Gly 829

Qy       2876 GACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTC 2935
              :::   :::         :::      |||:::   :::    |||   |||   |||
Db        830 AsnValAspGluThrGlyAsnAsnProSerValSerLeuLysLeuGluGluSerLysAla 849

Qy       2936 ACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGCAGAAACTGT 2995
              |||                                                  |||
Db        850 Thr---------------------------------------SerSerSerAsnPro 855

Qy       2996 TCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGT 3055
              |||:::|||   ||||||                 :::               ::::::
Db        856 SerSerProAlaProAspTrp------------TyrLysAspPheValThrAspAlaAsp 871

Qy       3056 GCAGAGACCGCTGACCATTCAGGGCAA 3082
              ||||||      :::|||||||||:::
Db        872 AlaGluValLeuGluHisSerGlyLys 880
```

RESULT 6
A89959
hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: A89959
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchima, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: A89959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1274 <KUR>
A;Cross-references: GB:BA000018; PID:g13701536; PIDN:BAB42830.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1562

```
Alignment Scores:
Pred. No.:              0.000287     Length:         1274
Score:                  174.00       Matches:        112
Percent Similarity:     35.28%       Conservative:   69
Best Local Similarity:  21.83%       Mismatches:     173
Query Match:            2.41%        Indels:         159
DB:                     2            Gaps:           20

US-09-302-812-3 (1-4069) x A89959 (1-1274)

Qy        311 AGGGTCCCACCGTCCTCGCCAGCC------------------------------------ 334
              |||:::|||   |||   |||:::
Db        203 ArgIleProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222

Qy        335 ---TGCGTCCCAGGGCAGGCGGGACAGCACAGA------------GGCAGCGCCACCTCG 379
                              |||      |||::::::
Db        223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGluGlnGluLysGlnAsnAsp 242

Qy        380 CTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCG 439
                 |||   ||||||   |||   :::          |||         |||
Db        243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerSerThrThr 262

Qy        440 GAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGT 499
              :::|||            ||||||   :::|||      :::||||||::::::   :::
Db        263 LysSerMetProAsnTyrSerLysValAspAsnThrIleLysIleGluAsnIleTyrAla 282

Qy        500 TCT-----------------------------------------GTACAAAAA------ 511
              |||                                         :::||||||
Db        283 SerGlnIleValGluGluIleArgArgGluArgGluArgLysValLeuGlnLysArgArg 302

Qy        511 ------------------------------------------------------------ 511

Db        303 PheLysLysAlaLeuGlnGlnLysArgGluGluHisLysAsnGluGluGlnAspAlaIle 322

Qy        512 ------------GATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAG 559
                          |||      |||               |||:::   |||      |||
Db        323 GlnArgAlaIleAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342

Qy        560 CTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACT 619
              |||:::   |||      ||||||:::   ||||||                  |||
Db        343 LeuAsnAspAspSerAspLeuThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362

Qy        620 GCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAA 679
                                                                      |||
Db        363 GlyIle---------------------------------------------------Glu 365

Qy        680 CCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCT 739
                 :::||||||:::               ::::::   |||            |||
Db        366 AsnGluThrValSer------------------AsnAspGluAsnLysGlnAlaSerIle 379

Qy        740 CAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTC 799
              ||||||:::|||   :::||||||      :::|||:::         |||   :::   :::
Db        380 GlnAsnGluAspThrAsnAspThrHisValAspGluSerProTyrAsnTyrGluGluVal 399

Qy        800 ACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAA------CACGCCAGA 853
              :::      :::::::::         ||||||   ::::::||||||         :::
Db        400 SerLeuAsnGlnValSerThrThrLysGlnLeuSerAspAspGluValThrValSerAsn 419

Qy        854 GAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGT 913
                       ||||||                                          |||
Db        420 ValThrSerGlnHisGln---------------------------------------Ser 426

Qy        914 TGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
                 |||      ::::::|||      |||      |||
Db        427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu------------------ 440

Qy        974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
                                ||||||   ::::::||||||               |||
Db        441 ------------------LysAsnGlnSerArgLeuIleAlaAspSerGluGluAspGly 454

Qy       1034 AATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCT 1093
                 :::               ||||||         |||   :::|||   |||
Db        455 AlaThr---------AsnLysGluTyrSerGlySerGlnIleAsp---AspAlaGlu 470

Qy       1094 AAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAA 1153
                          |||:::|||   ||||||:::|||:::         ::::::
Db        471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnIleGluAspAsn 490

Qy       1154 GAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGA------ 1207
                 :::   ::::::|||:::                  :::   |||
Db        491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluHisAlaVal 510

Qy       1208 ------------------------------------GACGCTGACATTGAATTTAGGAAA 1231
                                                  |||   :::||||||
Db        511 ThrGluSerGlnValAsnAsnIleAspLysThrValAspAsnGluIleGluLeuAlaPro 530

Qy       1232 CGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGT 1291
              |||:::      |||         :::      |||:::                  :::
Db        531 ArgHisLys---LysAspAspGlnThrAsnLeuAsnVal------------------Asn 543

Qy       1292 CGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATT---------TCTAGC 1342
                       ||||||:::|||               |||::::::         ||||||
Db        544 SerLeuLysThrAsnAspValAsn---------AspGlyHisValValGluAspSerSer 560

Qy       1343 CTGAAT------------GTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCT 1390
              :::|||            ||||||||||||         |||::::::
Db        561 MetAsnGluIleGluLysHis-AsnAlaGluIleThrGluAsnValGlnAsnGluAlaAl 580

Qy       1391 AAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGG 1450
                 :::|||:::                     |||      |||||||||   |||:::
Db        580 aGluSerGlu------------------------GlnAsnValGluGluLys-ThrIleG 592

Qy       1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCT 1510
```

```
                 |||          |||   :::   |||                |||     ||||||:
Db        592 luAsnValAsnProLysLysGlnThrGluLysValSerThrLeuSerLysArgProPheA 612

Qy       1511 CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGG------ 1564
              ::          |||       ||||||   ::::::   |||   |||   |||
Db        612 sn---------ValValMet-ThrProSerAspLysLysArgMetMetAspArgLysLys 628

Qy       1565 ---TGTGGGATCCGGCTGCCTCTCTTGAGACCA 1594
                    :::    :::|||    |||:::|||
Db        629 HisSerLysValAsnValProGluLeuLysPro 639
```

RESULT 7
D96730
unknown protein F5A18.20 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96730
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96730
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-884 <STO>
A;Cross-references: GB:AE005173; NID:g6453896; PIDN:AAF09079.1; GSPDB:GN00141
C;Genetics:
A;Gene: F5A18.20
A;Map position: 1

```
Alignment Scores:
Pred. No.:               0.000399        Length:          884
Score:                   171.50          Matches:         129
Percent Similarity:      34.48%          Conservative:    81
Best Local Similarity:   21.18%          Mismatches:      208
Query Match:             2.38%           Indels:          191
DB:                      2               Gaps:            29

US-09-302-812-3 (1-4069) x D96730 (1-884)

Qy         83 GGACGCGCAGTCCATCTCTCTCAGGTTAGTGAAATGAGGCTCTCCGCGGGGCCGGCCCGG 142
              |||       ||||||         |||          |||:::   :::   |||
Db        287 GlyGlyProValHisThrGluGln------HisMetGlnTyrAlaTyrGly--------- 301

Qy        143 GGACAGTGCGCTGCTGGTCCCAGC-----------ATGAATGCGGGCCCCGGCTGTGAA 190
                 |||   ||||||   ||||||                     ||||||
Db        302 AspGlnSerAlaAlaProProSerAsnPheSerAspHisAsnAla--------------- 316

Qy        191 CCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCG---------------CCG 235
                                    |||          |||||||||             |||
Db        317 --------------------TrpGlnProHisThrThrSerGlyValValTyrProPro 329

Qy        236 GCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTC---CTCGACCCCAAG 292
                 |||   |||:::|||   |||                 :::   :::   |||
Db        330 IleProSerSerAlaGlnSerIleProGlnHisAspSerSerMetAlaIleProProVal 349

Qy        293 GACGCTCACGTGCAG---------TTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTC--- 340
                 |||:::               |||   ||||||   :::|||      |||
Db        350 SerGlyHisIleMetProProTyrGlyArgPheProProProAsnProGlnProValGly 369

Qy        341 ---CCAGGGCAGGCGGGA-----------------------------------CAGCAC 361
                 |||         |||
Db        370 ProProTyrAlaPheGlyThrLysProProLeuHisProValAlaAlaPheMetAspAsp 389

Qy        362 AGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT 421
                 :::|||:::|||:::         |||:::         :::      :::|||:::
Db        390 SerTyrAlaAlaSerSerValProProLysLysAlaProValProAsnTrpLeuLysGlu 409

Qy        422 AAAGGAATCAAGACA-----------------------------------GCGGAATCA 445
              :::   :::|||                                          |||
Db        410 GluLeuLeuLysLysLysAlaAspLeuGlyArgProSerSerGlyArgPheGluGluArg 429

Qy        446 GAAAGTTTGGAT------------------------------------------------ 457
              ||||||:::|||
Db        430 GluSerMetAspAspAspValLeuTyrLysProProThrLysAlaAspGlnProAspLys 449

Qy        458 ---------AGTAAAGAAAACAAC------------------------------------ 472
                       |||         |||:::
Db        450 LysSerPheSerProSerAsnSerSerAspGluGluGluGluAspGluMetAspAlaAla 469

Qy        473 ---------------AATACAAGAATAGAATCCATGATGAGTTCTGTA---------CAA 508
                             |||         |||:::   ::::::::::         |||
Db        470 ArgThrThrGluIleAsnMetGluIleLysArgIleLeuThrGluValLeuLeuLysVal 489

Qy        509 AAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAG---CTAAGT 565
                 |||         :::               |||:::::::|||      :::   :::
Db        490 ThrAspGluLeuPheAspGluIleAlaThrLysValIleAsnGluAspGluAlaIlePro 509

Qy        566 CTTGATAAGTCACTCACTGAAAAA---------AGTACACAGTATTTGAACCAG------ 610
                 |||   |||:::                     |||:::      |||:::
Db        510 LysAspAspSerValGlnHisAsnHisLysLeuSerSerSerLeuLeuSerThrAlaAsp 529

Qy        611 ------CATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAG 664
                    |||:::   :::|||                  ||||||    :::|||:::
Db        530 ProLeuHisLysAlaSerAlaLysIleLeuValSerValGluGlyAlaAsnThrLysAla 549

Qy        665 CTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAAC 724
                    |||               :::   |||                   |||   |||:::
Db        550 SerSerGlySerProAlaAspValLeuGlyLeuAla------------SerTyrAlaSer 565

Qy        725 ATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT------ 778
                 |||                  |||      ||||||   :::::::::||||||
Db        566 AspAspAspAspAlaAspThrAspAlaAlaSerAspAlaAsnAlaAspGluAsnGlyVal 585

Qy        779 ---------------------AGAGACAAT---CAACAGTTTCTCACAACTGTAAAGCTT 814
                                   |||   |||      |||      :::|||   ||||||
Db        586 GluSerLeuGlyValGlySerArgHisAsnValSerGlnGlnProSerThrGluLysLeu 605

Qy        815 GCAAATGCAAAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAG 874
                                   :::   |||               ||||||
Db        606 ------------------ProAspProGluAlaMetAlaSerAlaLysLeuAspProAla 619

Qy        875 TGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGAT--------- 925
                 :::   :::   |||::::::         |||         :::|||
Db        620 ValGlyValAsnAlaAsnSerGlyLysAsnSerLysSerGlyLeuGluAspTyrSerGln 639

Qy        926 ------------------GAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTT--- 964
                                :::   |||            :::   :::|||||||||
Db        640 MetProGlySerThrArgLysAspAspGluAlaGlySerThrLysIleSerAspValSer 659

Qy        965 ---------GGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGA 1015
                       |||   :::|||      :::||||||:::      ::::::
Db        660 AlaSerSerGlyLeuAspAspAspThrSerGlySerArgLysGluHis------------ 675

Qy       1016 CAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCG 1075
                                         |||          :::|||:::      ::::::
Db        676 ---------------------------ProAspArgThrAspSerAspLysAspAlaIle 686

Qy       1076 ATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGT 1135
              :::|||         :::   |||||||||
Db        687 LeuAspGluProHisValLysAsnSer--------------------------------- 695

Qy       1136 CCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                       :::   |||   :::            |||   :::|||||||||
Db        696 ---------GlyValLysSerAspCysAsnLeuArgGlnAspSerAsnLysProTyrGly 712

Qy       1196 TTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGG---TACTCTACTAAGGGCGGT 1252
                       |||                  |||   |||         ||||||||||||
Db        713 LysAspLeuSerAspGluValSerThrAspArgSerArgIleValGluThrLysGlyGly 732
```

```
Qy       1253 GAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTA 1312
              :::   :::        |||        :::|||:::      |||    :::
Db        733 LysGluLysGlyAspSerGln---------AsnAspSerLysAspArgMetLysGluAsn 749

Qy       1313 AATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAA 1372
              :::   |||           :::      :::          :::::::::
Db        750 AspLeuLysSerAlaGluLysValLysGlyValGluSerAsnLysLysSerThrAspPro 769

Qy       1373 CATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGAC 1432
              |||   |||||||||||:::        :::        |||          ||||||
Db        770 HisValLysLysAspSerArgAspValGluArgProHisArgThrAsnSerLysGluAsp 789

Qy       1433 AGA------------------AGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA 1474
              |||                  ||||||:::     ::::::|||:::|||   |||
Db        790 ArgGlyLysArgLysGluLysGluLysGluGluGluArgSerArgHisArgArgAlaGlu 809

Qy       1475 AGGAAGATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACT 1534
                                                 |||  ||||||:::
Db        810 AsnSer--------------------------SerLysAspLysArg------------ 816

Qy       1535 CCCATTGAGGAGATGAGAAGAATGCCT 1561
                                  ||||||   |||
Db        817 ---------------ArgArgSerPro 820
```

RESULT 8
T18674
hypothetical protein T04F3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T18674; T24464
R;White, S.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19004
A;Accession: T18674
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3147 <WIL>
A;Cross-references: EMBL:Z74026; PIDN:CAA98419.1; GSPDB:GN00023; CESP:T04F3.1
A;Experimental source: clone B0240
R;Kershaw, J.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z19894
A;Accession: T24464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3147 <WI2>
A;Cross-references: EMBL:Z72513; PIDN:CAA96672.1; GSPDB:GN00023; CESP:T04F3.1
A;Experimental source: clone T04F3
C;Genetics:
A;Gene: CESP:T04F3.1
A;Map position: 5
A;Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/1; 2529/2; 2585/1; 2700/3; 2744/1; 2877/3;

```
Alignment Scores:
Pred. No.:               0.000956      Length:        3147
Score:                   168.00        Matches:       155
Percent Similarity:      33.67%        Conservative:  114
Best Local Similarity:   19.40%        Mismatches:    322
Query Match:             2.33%         Indels:        208
DB:                      2             Gaps:          30

US-09-302-812-3 (1-4069) x T18674 (1-3147)

Qy        356 CAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAA------------------------ 391
               :::||||||   ||||||   ::::::     :::
Db       1042 GluHisArgThrSerAlaValAsnIleAspLeuGluLysValPheIleHisGlySerSer 1061

Qy        392 ------------------------------CAAAAGACTATTACCAGTTGGATGGACACT 421
                                            ::::::  |||      :::       |||
Db       1062 LysLysProLysAsnAspAspGluLysIleArgArgGlyIleAlaGluPheGluArgThr 1081

Qy        422 AAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGA 481
              |||       |||   |||:::      ::::::               :::|||:::|||
Db       1082 LysGlnGluLysGluAlaGlnArgSerThrValIleGluThrSerGlnSerAsnSerArg 1101

Qy        482 ATA------GAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTA 535
              |||      ||||||   :::|||            |||        ||||||
Db       1102 IlePheGluGluSerSerIleSerMetAspAspValPheAsnAsnSerLeuHisAsnGlu 1121

Qy        536 GAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACA 595
                 ::::::      :::::::::   |||         |||         ||||||
Db       1122 SerGlnValSerGluIleThrGluAlaSerAspProSerAspLeuValLeuThrSerThr 1141

Qy        596 CAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACAC 655
                 :::   |||      :::               |||      ::::::
Db       1142 ThrPheHisAsnValIleGluGlu-----------LysIleAspAspAspValThrLys 1157

Qy        656 ACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGT 715
              |||:::      :::|||   |||   :::   |||   |||      ::::::|||
Db       1158 ThrAspSerAsnValGluGluGluLysGluGlnValArgLeuArgIleAspGluPheLys 1177

Qy        716 AATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAG 775
                       :::::::::   |||   :::                  ||||||
Db       1178 ArgProThrGluGluGlnAsnLeuGlnLysGluPheGluLeuThrLysLysGluGluGlu 1197

Qy        776 ---------AATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTT------------ 814
                       :::|||:::         ::::::   |||
Db       1198 TyrSerValLysMetGluAsnArgThrSerAlaValSerIleAspLeuAspLysValPhe 1217

Qy        815 ---GCAAATGCAAAGCAGACTACG------------------------------------ 835
                 ::::::|||:::||||||
Db       1218 AspGlnSerSerLysGluThrThrValSerAsnGluThrAspGluLysIleLysArgGly 1237

Qy        836 ------GAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAG------TGCAGCAAG 883
                    |||   |||   :::::::::   |||   |||:::         ::::::
Db       1238 IleAlaGluPheGluArgSerLysGlnGluLysGluValGlnArgSerGlyValAlaGlu 1257

Qy        884 TCTTGCCATCCTGGGGAA---------GACTGTGCAAGTTGTCAGCAAGATGAGATA--- 931
              :::   |||   |||:::         |||   ::::::         |||::::::
Db       1258 ThrSerHisSerGlyLysHisIlePheAspGluSerAsnIleSerMetAspAspValPhe 1277

Qy        932 ------------------------GACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGC 967
                                      |||         ||||||         |||
Db       1278 AsnThrSerGlnLysTyrLysSerAspGluLysLeuSerSerProGluArgThrValGlu 1297

Qy        968 TCTGAG---------------------------------------------GATGTTGGT 982
                 |||                                              :::
Db       1298 ProGluValSerThrAlaThrMetAsnLeuAspAsnIleIlePheAlaSerGlyIleAla 1317

Qy        983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
              |||      :::|||   :::   |||      :::|||
Db       1318 ThrArgGluGluAsnThrAspValLeuGluGluGluArgIleGlnLysArgValGlu 1337

Qy       1043 CCATTTGAGAAGGAAAGTGAACCC---GAATCACCGATGGATGTGGATAATTCTAAAAAT 1099
                 |||:::|||   :::|||      |||         :::|||      :::|||
Db       1338 GluPheLysLysThrThrGluAsnLeuGluIleGlnLysGluValValLeuThrLysGlu 1357

Qy       1100 AGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGAT 1159
                       :::|||:::      |||      :::         :::         |||
Db       1358 GluValAspAsnSerAspValLysGluHisArgThrSerAlaValAsnIleAspLeuAsp 1377

Qy       1160 GGT--------------AGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCA 1204
                               ||||||         |||                  :::
Db       1378 AspValPheIleGlnArgSerSerLysHisProGluAsnAspGluAspAspGluLysIle 1397

Qy       1205 AGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTA 1264
              |||         ||||||   :::               |||   |||   :::
Db       1398 ArgArgGlyIleAlaGluPheGluArgThrLysGlnGluLys------GluAlaGlnArg 1415

Qy       1265 CATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTA 1324
                       |||      :::|||         :::   |||
Db       1416 SerAlaValIleGluThrSerGlnSerAsnLysHisIlePheAsp--------------- 1430

Qy       1325 CCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAG 1384
                 |||||||||         |||      |||         |||:::|||:::|||
```

```
Db      1431 LysSerAsnIleSerMet------AspGluValPheAsnGluSerGlnAsnGlyGlnLys 1448

Qy      1385 GATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAA 1444
             ||||||        |||
Db      1449 AspSerSerAsnIleAsp------------------------------------------ 1454

Qy      1445 CAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCAC 1504
                         ||||||           ||||||        :::||||||       |||
Db      1455 MetLysGluThrAspMetProGluLysGluArgAspAspGlnArgTyrValAspValHis 1474

Qy      1505 CTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGG 1564
                      |||||||||          |||  |||              |||
Db      1475 ------ArgAspLysLys------------ProPheGluAsnGlyGluPheGluProThr 1488

Qy      1565 TGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTA 1624
                                     |||                  ||||||    :::
Db      1489 PheAsnGlySerLysIleSerAsnGluProLysGlnIleSerIleThrThrIleAsnLeu 1508

Qy      1625 GATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGG 1684
             |||                         |||||||||              |||
Db      1509 Asp------------------------AsnValPheProThrGluGluProLysLeuVal 1520

Qy      1685 GATAACAAGCATGTTAAAATGCCTTGTTCAGAACAA------------------------ 1720
                   :::   ::::::         |||:::
Db      1521 AlaGluAspAsnCysGluIleGluAlaGluGluGluArgIleArgLysArgIleLysGln 1540

Qy      1721 ------------------------------------AATTTGTACCCAGTGGAAGATGAG 1744
                                                 |||    |||   |||||||||
Db      1541 PheGluArgThrThrGlyGluGlnGluIleLeuLysAsnSerGluProAlaGluAspGlu 1560

Qy      1745 AATGGTGAG------------CGAACTGCGGGGAGCCGGTGGAGCTCATTCAGACTGCA 1792
                   :::             |||||||||            :::|||    :::
Db      1561 ThrSerAspGluLysLysHisArgThrAlaAlaValSerIleAspLeuAspLysValPhe 1580

Qy      1793 CTTCTCAACAAATTTACACGACCCCAAAAC------------------TTGAAGGATGCT 1834
             :::                 :::|||:::|||                   :::|||
Db      1581 ValGlnGlyThrAlaLysLysProGluAsnAspGluPheAspGluLysIleLysArgGly 1600

Qy      1835 ATTCTGAAATAC------------------------------------------AAT 1849
             |||   ::::::
Db      1601 IleAlaGluPheGluArgSerLysGlnGluLysGluValGlnArgSerGlyValAlaGlu 1620

Qy      1850 GTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTT 1909
                :::::::|||   |||      |||                  |||   |||
Db      1621 ThrSerHisSerSerLysHisIlePheAspGluSerAsnIleSerMetAspAspValPhe 1640

Qy      1910 GAAGAAGCAGAAGCTCAACATTTATATCAGTCC-------------------------- 1942
                   ::::::               |||:::|||
Db      1641 AsnThrSerGlnLys---------TyrLysSerAspGluLysLeuSerThrProGluArg 1657

Qy      1943 ---ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
                :::    |||::::::                 |||   ||||||
Db      1658 ThrValGluProGluValSerThrAlaThrMetAsnLeuAspAsnIleIlePheAlaSer 1677

Qy      2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
                      ::::::|||  |||       :::        :::|||:::|||
Db      1678 GlyIleAlaThrArgGluLys---AsnThrAspValLeuGluGluGluGluArgIleGln 1696

Qy      2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGA------------------ 2101
                   :::                        |||   :::
Db      1697 LysArgValGluGlu---------------PheLysLysThrThrGluAsnLeuGluIle 1711

Qy      2102 ------------------------------------AATGCTAAGATGAAATCGGAGTAT 2125
                                                 |||:::   :::|||
Db      1712 GlnLysGluValValLeuThrLysGluGluGlyAspAsnSerAspValLysAspHisLys 1731

Qy      2126 TCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCG 2185
             :::|||       :::|||:::   :::     :::|||      |||||||||:::   |||
Db      1732 AlaSerAlaValAsnIleAspLeuAspAspValPheIleGlnArgSerSerLysHisPro 1751

Qy      2186 ------------------GAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACA 2227
                                |||||||:::::            |||      |||
Db      1752 GluAsnAspGluAspAspGluLysIleArgArgGlyIleAlaGluPheGluArgThrLys 1771

Qy      2228 GAGAAAAAACCTACT------GGGTTGGTGACATTTACAAGACAGAGTCTTGAAGAT 2278
             ::::::|||                   |||         |||:::    |||     :::|||
Db      1772 GlnGluLysGluAlaGlnArgSerThrValIleGluThrGlnTyrSerSerLysAsp 1790
```

RESULT 9
T41581
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Aug-2003
C;Accession: T41581
R;Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z22002
A;Accession: T41581
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4717 <MUR>
A;Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08
A;Experimental source: strain 972h(-)
C;Genetics:
A;Gene: SPDB:SPCC737.08
A;Map position: 3
C;Superfamily: midasin (AAA ATPase with von Willebrand factor type A (vWA) domain)

```
Alignment Scores:
Pred. No.:               0.00124        Length:          4717
Score:                   167.00         Matches:         179
Percent Similarity:      34.63%         Conservative:    123
Best Local Similarity:   20.53%         Mismatches:      301
Query Match:             2.32%          Indels:          269
DB:                      2              Gaps:            46

US-09-302-812-3 (1-4069) x T41581 (1-4717)

Qy       434 ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATG 493
               ||||||            ||||||       |||:::               :::||||||
Db      3942 ThrAlaAsnGlnSerAspLeuAsp------GluSerGluAlaArgGluLeuGluSerAsp 3959

Qy       494 ATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAAT--------------------- 532
             |||:::   |||   ||||||:::            |||
Db      3960 MetAsnGlyValThrLysAspSerValValSerGluAsnGluAsnSerAspSerGluGlu 3979

Qy       533 ------------GTAGAAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTC 580
                         :::::::::   ||||||               |||      ||||||
Db      3980 GluAsnGlnAspLeuAspGluGluValAsnAspIleProGluAspLeuSerAsnSerLeu 3999

Qy       581 ACTGAAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAA 640
                ||||||                                                 |||
Db      4000 AsnGluLys------------------------------------------LeuTrpAsp 4005

Qy       641 AATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTA 700
                      :::       |||  |||||||||:::|||     ::::::
Db      4006 GluProAsnGlu------GluAspLeuLeuGluThrGluGlnLysSer---------Asn 4020

Qy       701 CCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCAC------ 754
                :::        :::|||          |||        :::     ||||||:::
Db      4021 GluGlnSerAlaAlaAsnAsnGluSerAspLeuValSerLysGluAspAspAsnLysAla 4040

Qy       755 ---AGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG 811
                |||   |||    :::|||       |||     ::::::             |||
Db      4041 LeuGluAspLysAspArgGlnGluLysGluAspGluGluMetSerAspAspValGly 4060

Qy       812 CTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAG 871
                                     :::||||||          :::   :::   :::
Db      4061 Ile---------------------AspAspGluIleGlnProAspIleGlnGluAsnAsn 4073

Qy       872 AAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGAT------ 925
                      |||:::            ||||||            :::|||
Db      4074 ------SerGlnProProProGluAsnGluAspHisLeuAspLeuProGluAspLeuLys 4091

Qy       926 ------------GAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
                         |||    ||||||         |||     |||    |||:::      |||
```

```
Db     4092 LeuAspGluLysGluGlyAspValSerLysAspSerAspLeuGluAspMetAspMetGlu 4111
Qy      974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
Db     4112 ---------------AlaAlaAspGluAsnLys---------GluGluAlaAspAlaGlu 4123
Qy     1034 AATTCTCCTCCA---------TTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTG 1084
Db     4124 LysAspGluProMetGlnAspPheGluAspProLeuGluGluAsnAsnThrLeuAspGlu 4143
Qy     1085 GATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAG------ACAAGTCCA 1138
Db     4144 AspIleGlnGlnAspAspPheSerAspLeuAlaGluAspAspGluLysMetAsnGluAsp 4163
Qy     1139 GGTTTTGATGAACAA------------------------GAAGATGGTAGTTCCTCCCAA 1174
Db     4164 GlyPheGluGluAsnValGlnGluAsnGluGluSerThrGluAspGlyValLysSerAsp 4183
Qy     1175 ACAGCA---------------------------------------AATAAACCT------ 1189
Db     4184 GluGluLeuGluGlnGlyGluValProGluAspGlnAlaIleAspAsnHisProLysMet 4203
Qy     1190 ---------TCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCT 1240
Db     4204 AspAlaLysSerThrPheAlaSerAlaGluAlaAspGluGlu---------------Asn 4218
Qy     1241 ACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTT------GAAGGAGGAGAGAGTCGC 1294
Db     4219 ThrAspLysGlyIleValGlyGluAsnGluGluLeuGlyGluGluAspGlyAlaAlaGlu 4238
Qy     1295 ACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAA 1354
Db     4239 SerGlyValArgGly---AsnGlyThrAlaAspGlyGluPheSerSerAlaGluGlnVal 4257
Qy     1355 TGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCT------------------------ 1390
Db     4258 GlnLysGlyGluAspThrSerThrProLysGluAlaMetSerGluAlaAspArgGlnTyr 4277
Qy     1391 ---AAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAG 1447
Db     4278 GlnSerLeuGlyAspHisLeuArgGluTrpGlnGlnAla------AsnArgIleHisGlu 4295
Qy     1448 TGG------------GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTT 1495
Db     4296 TrpGluAspLeuThrGluSerGlnSerGlnAlaPheAspAspSerGluPheMetHisVal 4315
Qy     1496 CCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACT------------CCCATTGAG 1543
Db     4316 LysGluAspGluGluGluAspLeuGlnAlaLeuGlyAsnAlaGluLysAspGlnIleLys 4335
Qy     1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db     4336 SerIleAspArgAspGluSerAlaAsn---------------GlnAsnProAspSerMet 4350
Qy     1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db     4351 AsnSerThrAsnIleAlaGluAsp------GluAlaAspGluValGlyAsp--------- 4365
Qy     1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db     4366 ---------LysGlnLeuGlnAspGly---------------------------GlnAsp 4373
Qy     1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db     4374 IleSerAspIleLys---GlnThrGlyGluAspThrLeuProThrGluPheGlySerIle 4392
Qy     1784 ---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTG 1840
Db     4393 AsnGlnSerGluLysValPheGluLeuSerGluAspGluAspIleGluAspGluLeuPro 4412
Qy     1841 AAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATC------------ 1888
Db     4413 AspTyrAsnVal---------LysIleThrAsnLeuProAlaAlaMetProIleAspGlu 4429

Qy     1889 ------GATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCC 1942
Db     4430 AlaArgAspLeuTrpAsnLys---HisGluAspSerThrLysGln--------------- 4443
Qy     1943 ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 2002
Db     4444 ---------------LeuSerIleGluLeuCys--------------------------- 4449
Qy     2003 CCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGT 2062
Db     4450 ---------------------------------------------GluGlnLeuArgLeu 4454
Qy     2063 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAG 2122
Db     4455 IleLeuGlu---------------------ProThrLeuAlaThrLysMetGlnGlyAsp 4467
Qy     2123 TATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGG--- 2179
Db     4468 PheArgThrGlyLysArgLeuAsnMetLysArgIleIleProTyrIleAlaSerGlnPhe 4487
Qy     2180 AAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCT 2239
Db     4488 LysLysAspLysIle---------------TrpMetArgArgValLysProSerLysArg 4502
Qy     2240 ACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGT 2299
Db     4503 ThrTyrGlnValMetIle---------SerIleAspAsp--------------------- 4512
Qy     2300 GAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAA 2359
Db     4513 ---------------------------------SerLysSerMetSerGluSerGlySer 4521
Qy     2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
Db     4522 ThrValLeuAlaLeuGluThrLeu---------------AlaLeuValThrLysAlaLeu 4536
Qy     2420 CTTGTGCAAGAA---------------GAAATCCGCTTTTTAATCAATCCTGAGTTGATT 2464
Db     4537 SerLeuLeuGluValGlyGlnIleAlaValMetLysPheGlyGluGlnProGluLeuLeu 4556
Qy     2465 ------ATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACA 2518
Db     4557 HisProPheAspLysGlnPheSerSerGlu--------------------------Ser 4567
Qy     2519 GGTACTGAGCAGTACAGTGAATACACA--------------------------------- 2545
Db     4568 GlyValGlnMetPheSerHisPheThrPheGluGlnSerAsnThrAsnValLeuAlaLeu 4587
Qy     2546 ------------------------GGCTATGCTGAGACATATCGTTGGTCCCGGAGCCAC 2581
Db     4588 AlaAspAlaSerMetLysCysPheAsnTyrAlaAsnThrAlaSerHisHisArgSerAsn 4607
Qy     2582 GAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
Db     4608 SerAsp------------------------IleArgGlnLeuGluIleIleIleSerAsp 4619
Qy     2642 GCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTG 2701
Db     4620 GlyIle-CysGluAspHisAspSerIleArgLysLeuLeuArgArg------AlaGlnGl 4637
Qy     2702 AACAAGGCTTACTGTGGATTTCTCCGTCCTGGAG 2735
Db     4637 uGluLysValMetIleValPheValIleLeuAsp 4648
```

RESULT 10
I38614
helicase II - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C;Accession: I38614
R;Stayton, C.L.; Dabovic, B.; Gulisano, M.; Gecz, J.; Broccoli, V.; Giovanazzi, S.; Boss
Hum. Mol. Genet. 3, 1957-1964, 1994

```
A;Title: Cloning and characterization of a new human Xq13 gene, encoding a putative heli
A;Reference number: I38614; MUID:95179111; PMID:7874112
A;Accession: I38614
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1641 <RES>
A;Cross-references: EMBL:U09820; NID:g606832; PIDN:AAC50069.1; PID:g606833
C;Genetics:
A;Gene: RAD54L

Alignment Scores:
Pred. No.:              0.00105     Length:        1641
Score:                  166.50      Matches:       178
Percent Similarity:     36.71%      Conservative:  134
Best Local Similarity:  20.94%      Mismatches:    322
Query Match:            2.31%       Indels:        216
DB:                     2           Gaps:          48

US-09-302-812-3 (1-4069) x I38614 (1-1641)

Qy     389 AAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448
           :::::::::  ::::::|||      :::||||||  |||:::::::     |||
Db     277 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSerSer 296

Qy     449 AGTTTGGATAGTAAAGAAAAC------AACAATACAAGAATAGAATCCATGATGAGTTCT 502
           |||   |||:::::::|||:::     :::|||   :::   :::        ||||||
Db     297 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysLysGlnArgThrSerSer 315

Qy     503 GTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA 562
              :::|||           :::                   :::     |||
Db     316 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 335

Qy     563 AGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCA 622
                 :::|||       |||:::       :::
Db     336 ArgLysGlnAlaAspIleThrSerSerSerSerSerAspIleGluAspAspAspGlnAsn 355

Qy     623 GCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTGGAAAGTGAACCT 682
           ::::::                 ||||||       ||||||   :::      :::|||
Db     356 SerIle------------GlyGluGlySerSerAspGluGlnLysIle------LysPro 369

Qy     683 CAAACAGTAACCCTGGTACCAGAGCAG---------TTTAGTAATGCTAACATTGATCGG 733
              |||     ||||||                  |||     ::::::   |||
Db     370 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 389

Qy     734 TCA----------------------CCTCAAAATGATGATCACAGTGACACAGATAGTGAA 772
           :::                         :::||||||     |||   |||   |||
Db     390 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAsnAspProGlu 409

Qy     773 GAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------CTTGCAAATGCAAAG 826
                 :::         |||          |||::::::   :::
Db     410 AsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuSerSerAspGlu 429

Qy     827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              :::   |||:::           |||   |||
Db     430 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 449

Qy     887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              :::||||||::::::    |||       |||     :::
Db     450 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer------------------ 463

Qy     947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGTCAAAAAATGAC------ 1000
                 ||||||       ||||||:::              ||||||
Db     464 ------GluSerAspSerAspSerGluGlu------------SerLysLysProArgTyr 477

Qy    1001 ---AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCA---TTTGAGAAG 1054
              ::::::|||:::|||     :::      :::      :::        ||||||
Db     478 ArgHisArgLeuLeuArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLys 497

Qy    1055 GAAAGTGAACCCGAATCACCGATGGATGTG---------GATAATTCTAAAAATAGTTGT 1105
           :::::::::|||:::       :::|||          :::      |||   |||
Db     498 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 517

Qy    1106 CAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGT 1165
           :::|||||||||      :::             |||      |||:::           |||
Db     518 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal------Ser 535

Qy    1166 TCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTT 1225
              |||:::         :::|||          |||:::   |||::::::|||
Db     536 GluSerGluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGlu 555

Qy    1226 AGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGA 1282
              :::|||   |||         |||         |||:::      |||   :::
Db     556 AsnGlnArgSerTyrLysGlnLysLysLysArgArgArgIleLysValGlnGluAspSer 575

Qy    1283 GGA--------------------------------------------------------- 1285

Db     576 SerSerGluAsnLysSerAsnSerGluGluGluGluGluGluLysGluGluGluGluGlu 595

Qy    1286 ------------------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCT 1327
                             |||         ||||||   ::::::|||   |||
Db     596 GluGluGluGluGluGluGluGluGluGluAspGluAsnAsp---AspSerLysSerPro 614

Qy    1328 GGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAA------ 1381
           |||                          :::      |||:::      |||
Db     615 Gly--------------------------LysGlyArgLysLysIleArgLysIleLeu 625

Qy    1382 AAGGATTCTAAAATC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
           ||||||   |||:::   |||:::               |||   :::   |||:::||||||
Db     626 LysAspAspLysLeuArgThrGluThrGlnAsnAlaLeuLysGluGluGluGluArgArg 645

Qy    1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA 1498
           |||         |||   :::   :::|||         |||:::   :::   :::
Db     646 LysArgIleAlaGluArgGluArgGluArg------GluLysLeuArgGluValIleGlu 663

Qy    1499 ---CCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT--------------- 1540
                    ||||||   |||      :::   |||   :::
Db     664 IleGluAspAlaSerProThrLysCysProIleThrThrLysLeuValLeuAspGluAsp 683

Qy    1541 GAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCC 1600
           ||||||   :::                             ||||||::::::
Db     684 GluGluThrLysGlu-----------------------ProLeuValGlnValHisArg 695

Qy    1601 AATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTT 1660
           |||      :::   :::                               ||||||
Db     696 AsnMetValIleLysLeu--------------------------------LysPro--- 703

Qy    1661 CCAACACATTATAAAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCT 1708
                 |||      |||          :::||||||
Db     704 ------HisGlnValAspGlyValGlnPheMetTrpAsp--------------------- 714

Qy    1709 TGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGC 1768
           |||      :::::::::             :::      ::::::::::   ||||||
Db     715 CysCysCysGluSerVal---------------LysLysThrLysLysSerProGlySer 729

Qy    1769 CGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAG 1828
                          |||         |||   |||                  |||:::
Db     730 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys---------------ThrLeuGln 744

Qy    1829 GATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTG 1885
                 :::   |||               |||   ||||||   |||||||||
Db     745 ValValSerPheLeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeu 764

Qy    1886 ATCGATTTCTGGGATAAGGTACTT---------------GAAGAAGCAGAAGCTCAACAT 1930
           :::            :::|||                  |||   |||
Db     765 ValGlyLeuSerSerSerIleLeuAlaPheAsnTrpMetAsnGluPheGluLys------ 782

Qy    1931 TTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGC 1990
              :::|||      |||   |||      |||:::            :::      :::
Db     783 ---TrpGlnGluGlyLeuLysAspAspGluLysLeuGlu------ValSerGluLeuAla 799

Qy    1991 ACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAA 2050
           |||         |||         |||:::   ::::::   :::         ||||||
Db     800 ThrValLysArgPro---------GlnGluArgSerTyrMetLeuGlnArgTrpGlnGlu 816

Qy    2051 CAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAG 2110
                 ::::::
```

```
Db        817 AspGlyGlyValMetIleIleGly---------------------------------- 824

Qy       2111 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGT 2170
                          |||      |||                |||   :::||||||
Db        825 ------------TyrGluMetTyr---------------ArgAsnLeuAlaGlnGlyArg 835

Qy       2171 TCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAG 2230
              :::      |||     ||||||||   :::|||          :::   :::   :::
Db        836 AsnVal---LysSerArgLysLeuLysGluIlePhe------AsnLysAlaLeuValAsp 852

Qy       2231 AAAAAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2290
                    |||      :::|||                |||::::::       |||
Db        853 ProGlyProAspPheValValCysAspGluGlyHisIleLeuLysAsn------GluAla 870

Qy       2291 GAAAGATGTGAAAAACCCTTG------------ACACGATTGCATGTCACTTACGAAGGT 2338
                       |||   :::              :::|||      :::           |||
Db        871 SerAlaValSerLysAlaMetAsnSerIleArgSerArgArgArgIleIleLeuThrGly 890

Qy       2339 ACC---------------------------------------------ATAGAAGAAAAT 2353
              |||                                             |||:::||||||
Db        891 ThrProLeuGlnAsnAsnLeuIleGluTyrHisCysMetValAsnPheIleLysGluAsn 910

Qy       2354 GGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTT------------GGAGGT 2401
                    |||   ::::::   :::|||   |||||||||:::                |||
Db        911 LeuLeuGlySerIleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGly 929

Qy       2402 GGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTG 2461
                       :::   :::|||         ::::::|||    :::
Db        930 GlnCysAlaAspSerThrMetVal------AspValArgValMetLysLysArgAlaHis 947

Qy       2462 ATTATTTCACGGCTC-----------------------------------TTCACTGAG 2485
              |||:::          :::                              |||:::
Db        948 IleLeuTyrGluMetLeuAlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLys 967

Qy       2486 GTGCTG------GATCACAATGAATGTCTAATTATCACAGGTACTGAG---CAGTACAGT 2536
                 |||             |||         |||     :::      |||      |||
Db        968 PheLeuProProLysHisGluTyrValLeuAlaValArgMetThrSerIleGlnCysLys 987

Qy       2537 GAATACACAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAA 2596
                 |||       ||| |||    |||:::    |||    |||              |||    ::
Db        988 LeuTyrGlnTyrTyr-LeuAspHisLeuThrGlyValGlyAsnAsnSerGluGlyGlyAr 1007

Qy       2597 AGGGACGACTGCGAGCGGCGCTGCACTG 2624
              :|||          :::||||||    |||
Db       1007 gGly------LysAlaGlyAlaLysLeu 1014
```

RESULT 11
T29043
hypothetical protein B0228.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29043
R;Leimbach, D.
submitted to the EMBL Data Library, March 1995
A;Description: The sequence of C. elegans cosmid B0228.
A;Reference number: Z18324
A;Accession: T29043
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1788 <LEI>
A;Cross-references: EMBL:U23168; PIDN:AAC38806.1; CESP:B0228.2
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:B0228.2
A;Introns: 1456/2; 1482/3; 1516/2; 1551/3; 1595/3; 1646/1; 1671/1; 1716/2; 1749/3

Alignment Scores:

Pred. No.:	0.00267	Length:	1788
Score:	161.00	Matches:	213
Percent Similarity:	31.45%	Conservative:	126
Best Local Similarity:	19.76%	Mismatches:	411
Query Match:	2.23%	Indels:	328
DB:	2	Gaps:	45

```
US-09-302-812-3 (1-4069) x T29043 (1-1788)

Qy        194 TGCACCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGG 253
              ||||||||                   |||:::|||::::::    ||||||      |||
Db        808 CysThrLysGlu------------AlaSerThrSerAlaAlaAlaAlaLeuSerAlaGly 823

Qy        254 AGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGG 313
              |||
Db        824 Ser--------------------------------------------------------- 824

Qy        314 GTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCC 373
                          |||   :::|||             :::         :::   |||
Db        825 ---------LysSerGluSerCysLysIleValArgLeuAlaSerAsnLeuGlyHisPro 841

Qy        374 ACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAG 433
              |||||||||||
Db        842 ThrSerLeuValLeu--------------------------------------------- 846

Qy        434 ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATG 493
                          ||||||         :::|||   ||||||
Db        847 ---------CysGluSerSerSerValGlnGluThrAsnAsnValHis------------ 859

Qy        494 ATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAA--------------- 538
                          |||:::|||        ::::::|||      |||
Db        860 ---------TyrGlnArgAspGluHisHisGluHisIleSerGluThrLysSerTyrPro 876

Qy        539 -----------------------------AAATTAGTAAATGTTTCTCAGCTAAGTCTT 568
                                          |||          :::::::::   :::
Db        877 ArgAspGlyGlyLysPheThrLeuAspThrLysAlaSerThrAlaAsnGluValArgIle 896

Qy        569 GATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATG 628
              ||||||   |||      :::||||||    :::      |||            :::|||
Db        897 AspLysAspLeuGluLysLysSerAspArgGluLeu---------GluThrGluIleLys 913

Qy        629 TGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACA 688
                    :::|||||||||:::         |||         :::   :::         ::::::
Db        914 ThrIleValArgAsnGluGlyGluProValGluIlePheValSerAlaThrGluGluSer 933

Qy        689 GTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAAC--------------------ATT 727
                       |||           |||   ||||||                         :::
Db        934 AlaAlaGlyValThrThrSerLeuSerArgAlaAsnProPheGluSerAlaAsnIleLeu 953

Qy        728 GATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAAT 787
                    ||||||         :::   :::   :::         |||:::   :::
Db        954 LeuThrSerProAsnLysGlyGluProAlaTyrSerArgValThrGluSerSerGlu--- 972

Qy        788 CAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAACAC 847
                       ||||||                                   |||   :::
Db        973 ---------LeuThr-------------------------GluThrAsnAsnValGln 980

Qy        848 GCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGT 907
                 |||      :::   ||||||:::      |||
Db        981 LeuArgArgGluGluGluHisGlnGluThrGluLysIle--------------------- 993

Qy        908 GCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGC 967
                                  |||::::::                  ::::::   |||
Db        994 ---------------------IleGluIleAla---------------AlaAsnGlyGly 1001

Qy        968 TCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGC 1027
              |||      :::      |||         |||   |||            |||:::
Db       1002 SerSerLeuLeuArgAlaGlyPheAla---AspGluLysPheAlaAspIleGluAlaLys 1020

Qy       1028 CTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGAT 1087
              ||||||         ||||||                  ::::::      :::
Db       1021 LeuGlyLysAspAlaGlnPheGlu---------SerAlaGlnThrIleArgGlnIleGly 1037

Qy       1088 AATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGAT 1147
              |||         :::                  |||   :::|||||||||      ||||||
Db       1038 AsnGluAspLysThrAsnLeuSerIleGlyAlaSerGlnGluThrSerValThrPheAsp 1057

Qy       1148 GAA------CAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAA 1201
              |||      |||   :::   ||||||   :::         :::                   :::
```

```
Db       1058 GluThrValGlnCysAsnLysSerSerCysGluGluThrSerIleThrLysValAlaLys 1077

Qy       1202 GCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAA------ 1255
                   :::     :::   ||||||           ||||||:::       :::
Db       1078 AsnIleGluProHisValIlePheArg---------SerThrGluAlaSerAspMetAla 1094

Qy       1256 GTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAAT 1315
              |||   :::|||:::                 |||                 ::::::
Db       1095 ValGlyIleHisTyrThrLeuArgSerSerGluLysValGluGluThrGluGluIleLys 1114

Qy       1316 GCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACAT 1375
                 :::      |||      |||            |||     |||
Db       1115 ---AsnIleAlaArgAsnGlyGlySerAlaThrPheSerCysPheAlaSerGlyAsp--- 1132

Qy       1376 GGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGA 1435
                 :::         ::::::      |||   |||   |||:::   |||
Db       1133 ---GluSerProAspSerValSerAlaPheLeuThrArgGlnProGlnGluGlu------ 1149

Qy       1436 AGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTT 1495
                                                ||||||            |||
Db       1150 ------------------------------ThrThrGlu------------LysLeuPhe 1155

Qy       1496 CCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGA 1555
              |||       :::            |||:::         |||      ||||||
Db       1156 ProThrProMetPheAspPheIleLysPheAsnSerThrAlaAlaGluGluPheAlaVal 1175

Qy       1556 ATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACT 1615
                                                         |||   |||:::
Db       1176 Trp---------------------------------------AsnThrThrIlePhe 1181

Qy       1616 ATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAA 1675
                 |||   |||               ||||||||   |||   |||   |||
Db       1182 ArgArgLysAsp------AsnGluGlyGluValGluLysIlePheAsnThrSer------ 1197

Qy       1676 GATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTG 1735

Db       1197 ------------------------------------------------------------ 1197

Qy       1736 GAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTT 1795
              |||      :::   |||         :::   :::      |||   :::         |||
Db       1198 GluAlaGlyHisAsnGluThrPheSerAlaAsnAlaAlaGluAspValSerValThrLeu 1217

Qy       1796 CTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCA 1855
                                                ||||||   |||      :::   |||
Db       1218 --------------------------------AspAlaAspLeuHisPheGlyValGly 1226

Qy       1856 TATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA--------- 1906
              |||   :::            |||                 |||:::
Db       1227 TyrLysGluHisArgGlnIleThrLys---------------AspGluAlaAsnGlnGly 1241

Qy       1907 ------------CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                          :::         ::::::          ::::::         |||
Db       1242 GluGlyThrGlyMetHisSerGlyAlaSerGluGluThrIlePheAsnLeuGlyTyrAsp 1261

Qy       1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
                     |||                :::|||                  ||||||
Db       1262 TyrCysLysGlnProThrGluPheThrThrValCysValThrGluAspLysLeuLeuIle 1281

Qy       2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
              |||         :::         ::::::|||:::         :::|||   |||:::
Db       1282 GlnGlyAla---TyrGlyPheArgAlaAlaLysGluGluSerIleThrLeuAspAlaAsp 1300

Qy       2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
                 |||         |||         |||         |||
Db       1301 LeuHisPheGlyValAsnTyrArgAspLeuLeuAlaMetGlyLeuHisAlaSer------ 1318

Qy       2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                             |||      ||||||         |||
Db       1319 ---------------AsnAsnGluIleGluGlyThrGlyMetArgSerThrAlaSerGlu 1333

Qy       2195 AAAACGCTCTTC------------------TGCTACTTTAGAAGAGTCACAGAGAAAAAA 2236
              :::|||:::|||                  |||                     |||:::
Db       1334 GluThrIlePheAsnLeuAlaTyrAspTyrCys---------------------LysGln 1346

Qy       2237 CCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGA 2296
              ||||||           |||                     |||
Db       1347 ProThrGluPheArgThrValPheValSerGluAspHisGlnPheValHisGlyAlaPhe 1366

Qy       2297 TGTGAAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGC 2356
                 :::   :::               |||:::         |||      :::   |||
Db       1367 GlyPheArgAlaValGlyGluGluHisIle------GluThrGlnValLeuGlu------ 1382

Qy       2357 CAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCA 2416
                       ||||||            |||   |||   |||   :::
Db       1383 ---------LeuGlnAlaArgMetValGluValMetValGluGlySerValHisAsnLeu 1399

Qy       2417 GGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTC 2476
                       |||:::   |||:::                                  |||
Db       1400 AlaArgArgHisGluAspGluProPheVal---------------------------Leu 1410

Qy       2477 TTCACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGT 2536
              :::|||||||||::::::                  ||||||      ||||||
Db       1411 TyrThrGluValIleGluGluThr---------IleIleArgValAspGluGlnLeuGlu 1427

Qy       2537 GAATACACAGGCTATGCTGAGACATATCGTTGGTCC------CGGAGCCACGAAGATGGG 2590
              :::   |||         ||||||   :::   |||         :::      |||   |||
Db       1428 LysLysThrThrValIleGluThrGluGlnAlaSerGluValLysMetArgGluLysGly 1447

Qy       2591 AGTGAAAGGGACGACTGCGAGCGGCGCTGCACT---------------GAGATCGTTGCC 2635
                 ||||||         |||:::|||   :::                 :::
Db       1448 GluGluArgArgLysGluGluLysArgValSerPheAlaAlaGluValGlnGluLysThr 1467

Qy       2636 ATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCT--------------- 2680
              ::::::|||:::               ||||||      :::
Db       1468 MetGluAlaIleAspLysSerLeuGlyLeuAspThrSerMetGluValGluProAlaPhe 1487

Qy       2681 ---------------------------------------------GAGAAAATGAGACGCGAG 2698
                                                           |||:::      |||:::
Db       1488 GlnLysProSerIleIleLysLysProMetLysLysGluArgGluArgArgSerArgAsp 1507

Qy       2699 CTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAGAATCTTTCTGCA 2758
              |||   :::               :::|||         :::      |||   |||
Db       1508 LeuArgGlnAsnAlaAlaProAlaPheLysProValArgArgAsnSerLeuLeuGlnAla 1527

Qy       2759 GTGGCCACAGGA------------------------------------------------ 2770
              :::|||   |||
Db       1528 LeuAlaIleGlySerProHisAsnIleProHisPheLysThrLeuAspAspIleValLys 1547

Qy       2770 ------------------------------------------------------------ 2770

Db       1548 AlaIleLysHisAlaGlyLeuGluTyrSerAsnLeuIlePheGlyIleAspTyrThrLys 1567

Qy       2771 ---AACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG--------------------- 2806
                 |||:::      |||                  |||   |||
Db       1568 SerAsnPheTyrGlnGlyGluArgThrPheAspLysArgProLeuHisThrIleAspPro 1587

Qy       2807 ---------------TTAAAAGCCTTAATACAGATA------------TTGGCAGCTGCT 2839
                             :::   :::|||||||||               |||::::::
Db       1588 AlaGluMetAsnProTyrGlnGlnValIleGlnIleValGlyLysThrLeuSerSerPhe 1607

Qy       2840 GCAGCTGAGCGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATT 2899
                 |||:::      :::      :::   |||||||||   |||
Db       1608 AspAlaAspGlyGlnIleProAlaTyrGlyPheGlyAspGluGlu--------------- 1622

Qy       2900 TACAGCATGCAC---ATTTTC---CTTACTGAAAGGAAACTCACTGTTGGAGATGTG--- 2950
              ::::::   |||   ||||||   :::   ||||||               |||
Db       1623 PheThrAspHisGlyIlePheAsnIleAlaGluArgTyrAspLeuGluLysAspCysAsn 1642

Qy       2951 ---TATAAGCTGTTGCTACGATACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGA 3007
                 ::::::   :::||||||   |||||||||                        |||
Db       1643 GlyPheGluGluValLeuArgValTyrAsnGluValThrProThrIleGluMetSerGly 1662

Qy       3008 CCAGACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAG 3061
              |||            |||   |||      |||:::|||   |||   |||
Db       1663 Pro---ThrAsnPheValProLeuIleAspArgAlaIleGluIleCysLysGlu 1679
```

```
RESULT 12
T24806
hypothetical protein T10G3.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24806
R;Burton, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19937
A;Accession: T24806
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1164 <WIL>
A;Cross-references: EMBL:Z81118; PIDN:CAB03330.1; GSPDB:GN00023; CESP:T10G3.5
A;Experimental source: clone T10G3
C;Genetics:
A;Gene: CESP:T10G3.5
A;Map position: 5
A;Introns: 15/3; 73/3; 387/3; 412/3; 708/3; 818/3; 894/3; 982/3; 1080/2

Alignment Scores:
Pred. No.:               0.00285     Length:         1164
Score:                   160.00      Matches:        168
Percent Similarity:      36.41%      Conservative:   171
Best Local Similarity:   18.05%      Mismatches:     360
Query Match:             2.22%       Indels:         232
DB:                      2           Gaps:           33

US-09-302-812-3 (1-4069) x T24806 (1-1164)

Qy    374 ACCTCGCTTGTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAG 433
Db    253 ThrGluLeuValAsnAlaGlnLysLeu------------MetAspAla------------ 264

Qy    434 ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATG 493
Db    265 IleSerGlnGluLysAspIleGluIleLysGluHisLeuAsnSer---IleArgAsnLeu 283

Qy    494 ATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTA------ 547
Db    284 SerMetGluArgGluLysGlnHisIleValAsnGluAsnLeuGluLysLysIleGlyGlu 303

Qy    548 ---------AATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAG 598
Db    304 GlyGluGluThrValLysGlnLeuGlnIleSerTyrAspAlaGlnSerGluGluLeuLys 323

Qy    599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
Db    324 GlnArgAsnGluArgValVal------------------GlnLeuGluAlaArgIleGlu 337

Qy    659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
Db    338 GluAsnValPheGlu--------------------------------------------- 342

Qy    719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
Db    343 ------LeuSerGluAsnLysGlnAsnValLysArgLeuGluAspLysValGlnGluSer 360

Qy    779 AGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAG--------- 829
Db    361 GlnAspAlaLeuGlnMetLeuSerAsnIleAsnGlySerAsnGluGluGlnMetIleSer 380

Qy    830 ---------------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCAC 868
Db    381 LeuAsnSerLysPheGluArgAsnThrAlaGluArgLysArgIleGluAlaValPheGlu 400

Qy    869 CAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAG 928
Db    401 GluLysValThrValGln---------GlyGluArgLeuLysThrLeuGluMetAlaAsn 417

Qy    929 ATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCT---------GAGGATGTT 979
Db    418 LeuAspLeuThr------AsnGluLeuAlaSerMetGlySerLeuLeuAspLysGluArg 435

Qy    980 GGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCT 1039
Db    436 SerLeuLeuGluGluLysAsnLysGluIleSerGluArgAspSerSerIleAsnAspLeu 455

Qy   1040 CCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAAT 1099
Db    456 ------------LysGluLysLeuAlaGluSerGluLysLysAlaThrLysTyrLysAsn 471

Qy   1100 AGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGAT 1159
Db    472 Glu-LeuLysGluHisAlaAspLeuValGluAsnLeuThrLeuGlnLeuAsnLysLeuGl 491

Qy   1160 GGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATT 1219
Db    491 n-----------------------------GluAsnSerLysAspLeuMetGluLysIl 501

Qy   1220 GAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAA 1279
Db    501 eSerAlaGlyGluGlyGlyAlaLysMetAla---------------IleGluGlnLeuGl 516

Qy   1280 GGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCT 1339
Db    516 uGlnGluLysValLysLeu--ThrAsnGluLeuGlnThrSerSerGlu------------ 531

Qy   1340 AGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAAATCACA 1399
Db    532 ------------------LysThrLysLysAlaSerGlyGluLeuGluAlaLysIleSer 545

Qy   1400 GATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA------AAAGAACAG---TGG 1450
Db    546 GluLeuGluLysLysLeuArgAspAlaGluAlaSerArgThrAspLysGluGlnLysTrp 565

Qy   1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCT 1510
Db    566 LysGlnGluLysGluSerPheGluArgLys------------------------------ 575

Qy   1511 CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTGTGGG 1570
Db    576 ---------------LeuAlaGluAlaGluAspGluIleLysArgLysGlyGluArgPhe 590

Qy   1571 ATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTT 1630
Db    591 ValGluMetGluLysGluMetGluGluGluArgGlnLysAlaThrAspArgThrLeuLys 610

Qy   1631 TTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAAC 1690
Db    611 Leu------------------------------------LysAspAlaLeuValAsn 617

Qy   1691 AAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGT 1750
Db    618 ---------------------SerGluLysAsnLeuGluThrIleLysLysGluSerGlu 630

Qy   1751 GAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACA 1810
Db    631 AspArgGlu---------------LysIleValArgGluLysAspAlaHisLeuGluGlu 645

Qy   1811 CGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGG 1870
Db    646 AsnLysLysArgIleGluAspAlaValGlnLysLeuGluGluAlaGluLysArgAlaArg 665

Qy   1871 GACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACAT 1930
Db    666 GluLeuGluAlaSerValSerSerArgAspThrThrValSerThrLysGluSerGlu--- 684

Qy   1931 TTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGC 1990
Db    685 ---------------LeuSerGluLeu---LysGlyLysLeuThrGluSerAsnSerPhe 698

Qy   1991 ACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAA 2050
Db    699 IleGluGluLeuLysValGlnValGluLysValSerAsnGluIleSerGluLysGlnGln 718
```

```
Qy     2051 CAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAG 2110
            ::::::   :::|||:::|||                              :::   |||
Db      719 GluValGluAsnLeuMetAlaGlu---------------MetArgAspLysGluAlaHis 733

Qy     2111 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGT 2170
               |||::::::                                  ||||||   :::
Db      734 TrpLysThrLysArgAspGlu---------------------------PheGluAlaGln 744

Qy     2171 TCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAG 2230
                  ||| |||:::   :::   :::   :::|||:::||||||:::
Db      745 MetLeuArg-AsnGlnGluAspAsnGluGluAlaSerSerThrLeuLysSer-------- 761

Qy     2231 AAAAAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2290
                                         :::|||::::::::::|||     :::
Db      762 -------------------------ValGlnGluGlnLeuMetLysGluLysGluThrSe 773

Qy     2291 GAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAA 2350
               :::   ||||||
Db      773 rGlyGluGluLysAsnGln------------------------------------------ 779

Qy     2351 AATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACC 2410
                                                   |||:::   |||
Db      780 ---------------------------------------LeuIleSerValLysSerGl 786

Qy     2411 AGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCA 2470
             :::::::::|||  :::   |||:::      ||||||    |||
Db      786 nLeuGluGluLeu--LysThrGluValGluArgLeuIleArgSerGlu------------ 801

Qy     2471 CGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAG 2530
                     |||:::  ::::::    :::    :::      |||::::::
Db      802 GluGluLysThrGlnGluIleGluLysLeuLysSerAlaValThrAlaThrThrGlnGlu 821

Qy     2531 TACAGTGAATACACAGGCTATGCTGAGACATATCGTTGG--------------------- 2569
              |||   |||     :::|||:::   |||
Db      822 ArgAspGluLeuThrAlaThrSerGluSerLeuArgThrGluCysGluAsnLeuAsnSer 841

Qy     2570 ---------------------TCCCGGAGCCACGAAGATGGAGTGAAAGGGACGACTGC 2608
                                 ||||||   |||   :::       ::::::
Db      842 LysIleGlnSerIleGluGluSerArgArgHisAlaGluGluLysGlySerGluAsnLeu 861

Qy     2609 GAGCGGCGCTGCACTGAGATCGTTGCCATCGATGCT---CTTCACTTCAGACGCTACCTC 2665
            ||||||      ||||||         ::::::    :::     |||
Db      862 GluArgMetIleThrGluLysSerArgLeuGluLysAspIleGluGluArgGluSerThr 881

Qy     2666 GATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTC 2725
               |||   :::   |||   :::   :::   |||:::
Db      882 IleGlnSerIleGlnGluAlaLeuGluThrLysAspAsnGluIleGluSerLeuLysThr 901

Qy     2726 CGTCCTGGAGTTTCTTCAGAGAATCTT------------TCTGCAGTGGCCACAGGAAAC 2773
                  |||      :::   |||              |||   :::   :::   |||
Db      902 ThrGlnArgValValGluAspGluLeuValSerLysIleSerHisIleGluSerPheAsn 921

Qy     2774 TGGGGCTGTGGTGCCTTTGGGGGTGAT---GCCAGGTTAAAAGCCTTAATACAGATATTG 2830
                        |||      :::   |||      |||      |||:::   |||
Db      922 SerArgIleGluGluPheGluLysGluMetAlaSerGlyLysArgThrIleGluArgLeu 941

Qy     2831 GCAGCTGCTGCAGCTGAG------CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAA 2884
             |||      ||||||            :::|||   ||||||       |||:::
Db      942 GluAlaGluLysAlaGluGluThrGluLysLeuValValPheThrGlyThrGlnSerGln 961

Qy     2885 TTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACTGTT--- 2941
                   ::::::     :::      :::   |||::::::   |||:::
Db      962 LysGlnGluGluLeuGluLysLeuGlnLysGluIleGlnGluLysGluThrThrIleAla 981

Qy     2942 ------------------------------------------GGAGATGTGTATAAGCTGTTG 2962
                                                ||||||   :::   |||
Db      982 ArgMetThrSerSerLysThrGlnPheGluAlaMetPheAlaAspValGlnGlnThrLeu 1001

Qy     2963 CTACGATACTACAATGAAGAA 2983
               :::      |||::::::
Db     1002 SerLysGluIleAsnAspLys 1008
```

RESULT 13

```
T38077
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38077
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z21767
A;Accession: T38077
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1957 <CON>
A;Cross-references: EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SPAC1F3.06c
A;Experimental source: strain 972h-; cosmid c1F3
C;Genetics:
A;Gene: SPDB:SPAC1F3.06c
A;Map position: 1

Alignment Scores:
Pred. No.:              0.00321       Length:         1957
Score:                  160.00        Matches:        182
Percent Similarity:     35.63%        Conservative:   149
Best Local Similarity:  19.59%        Mismatches:     343
Query Match:            2.22%         Indels:         255
DB:                     2             Gaps:           39

US-09-302-812-3 (1-4069) x T38077 (1-1957)

Qy      242 TCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCAC 301
              |||      |||      |||   ||||||   :::|||:::   |||      |||
Db      508 SerThrGlnSerSerLeuGlnSerLeuGlnArgAspIleLeuAsnGluLysLysLysHis 527

Qy      302 GTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCAC 361
                  :::                                      |||      :::
Db      528 GluValTyrGlu------------------------------SerGlnLeuAsnGluLeu 537

Qy      362 AGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT 421
            :::|||      :::   :::         :::   ::::::|||   :::   |||
Db      538 LysGlyGluLeuGlnThrGluIleSerAsnSerGluHisLeuSerSerGlnLeuSerThr 557

Qy      422 AAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGA 481
                         ||||||   |||:::            ||||||
Db      558 ---------LeuAlaAlaGluLysGluAla------AlaValAlaThrAsnAsnGluLeu 572

Qy      482 ATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAA 541
               ||||||   |||                  |||   :::|||      :::   |||
Db      573 SerGluSerLysAsnSerLeuGlnThrLeuCysAsnAlaPheGlnGluLysLeuAlaLys 592

Qy      542 TTAGTAAATGTTTCTCAGCTA---------------------AGTCTTGAT------ 571
                  :::|||   ||||||                        |||||||||
Db      593 ------SerValMetGlnLeuLysGluAsnGluGlnAsnPheSerSerLeuAspThrSer 610

Qy      572 ---AAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATG 628
             |||   |||   |||         :::   |||   |||||||||
Db      611 PheLysLysLeuAsnGluSerHisGlnGluLeuGluAsnAsnHisGlnThrIle------ 628

Qy      629 TGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACA 688
                                       |||:::|||||||   ::::::      :::
Db      629 --------------------------ThrLysGlnLeuLysAspThrSerSerLysLeu 639

Qy      689 GTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGAT 748
                |||         |||      ||||||   ::::::
Db      640 GlnGlnLeu---------GlnLeuGluArgAlaAsnPheGluGlnLys------------ 652

Qy      749 GATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTA 808
                  |||   |||:::||||||   |||   :::      |||         :::
Db      653 ------GluSerThrLeuSerAspGluAsnAsnAspLeuArgThrLysLeuLeuLysLeu 670

Qy      809 AAGCTTGCAAATGCAAAG------------------------------------------ 826
            :::   :::|||
Db      671 GluGluSerAsnLysSerLeuIleLysLysGlnGluAspValAspSerLeuGluLysAsn 690

Qy      827 ---CAGACTACGGAAGATGAACACGCCAGA------------------------------ 853
```

```
                 ||||||  :::::::::     :::
Db   691 IleGlnThrLeuLysGluAspLeuArgLysSerGluGluAlaLeuArgPheSerLysLeu 710

Qy   854 GAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGT 913
         ||||||||:::   ::::::      :::           :::
Db   711 GluAlaLysAsnLeuArgGluValIleAspAsnLeuLysGlyLysHisGluThrLeuGlu 730

Qy   914 TGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
            |||:::::::::::::               |||  |||||||||
Db   731 AlaGlnArgAsnAspLeu------------HisSerSerLeuSerAsp------------ 742

Qy   974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
                        :::||||||  |||  ::::::    |||
Db   743 ---------------AlaLysAsnThrAsnAlaIleLeuSerSerGlu------------ 753

Qy  1034 AATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCT 1093
                     |||  ||||||          :::        |||
Db   754 ------------LeuThrLysSerSerGluAspValLysArgLeuThrAla---AsnVal 768

Qy  1094 AAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAA 1153
         :::         |||||||||:::||| |||::::::          |||:::|||
Db   769 GluThrLeuThrGlnAspSerLysAla-MetLysGlnSerPheThrSerLeuValAsnSe 788

Qy  1154 GAAGATGGTAGTTCCT------------------CCCAAACAGCAAATAAACCTTCAAGG 1195
                  :::                         :::|||:::|||
Db   788 rTyrGlnSerIleSerAsnLeuTyrHisGluLeuArgAspAspHisValAsnMetGln-- 807

Qy  1196 TTCCAAGCAAGAGACGCTG-----------------------------------ACATT 1219
          |||:::     |||||                                   ||
Db   808 -SerGlnAsnAsnThrLeuLeuGluSerGluSerLysLeuLysThrAspCysGluAsnLe 827

Qy  1220 GAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAA 1279
         |     |||  ||||||:::       |||:::  :::|||     :::
Db   827 uThrGlnGlnAsnMetThrLeuIleAspAsnValGln--LysLeuMetHisLysHisVal 846

Qy  1280 GGAGGAGAGAGTCGCACTGGA---ATGAATGATTTAAATGCTAAACTACCTGGAAATATT 1336
               ||||||:::      :::  ::::::|||  ||||||      ::::::
Db   847 AsnGlnGluSerLysValSerGluLeuLysGluValAsnGlyLysLeuSerLeuAspLeu 866

Qy  1337 ------------TCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAG 1384
                     |||||||||||||||          ::::::  |||
Db   867 LysAsnLeuArgSerSerLeuAsnValAlaIleSerAspAsnAspGln------------ 882

Qy  1385 GATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAA 1444
                   |||        |||     |||  |||      |||       ::::::
Db   883 ---------IleLeuThrGlnLeuAlaGluLeuSerLysAsnTyrAspSerLeuGluGln 899

Qy  1445 CAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCAC 1504
               |||:::         :::    |||
Db   900 ------GluSerAlaGlnLeuAsnSerGlyLeuLys---------------------Ser 910

Qy  1505 CTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGG 1564
         |||     :::|||:::   |||   |||      ||||||:::
Db   911 LeuGluAlaGluLysGlnLeuLeuHisThrGluAsnGluGluLeuHis------------ 926

Qy  1565 TGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTA 1624
               |||||||||   |||                     |||    :::::::::
Db   927 ------IleArgLeuAspLysLeu------------------ThrGlyLysLeuLysIle 938

Qy  1625 GATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGG 1684
         :::     ::::::  ::::::     |||                 :::::::::
Db   939 GluGluSerLysSerSerAspLeuGlyLysLysLeuThrAlaArgGlnGluGluIleSer 958

Qy  1685 GATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAG 1744
         :::  |||      |||      |||  |||   :::   |||:::   :::
Db   959 AsnLeuLysGluGluAsnMetSerGlnSer---GlnAlaIleThrSerValLysSerLys 977

Qy  1745 AATGGTGAGCGA---ACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAAC 1801
                |||      :::      |||:::  |||                |||   |||
Db   978 LeuAspGluThrLeuSerLysSerSerLysLeuGluAlaAspIleGluHisLeuLysAsn 997

Qy  1802 AAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCT 1861
         |||   :::      :::      ::::::|||:::|||      |||
Db   998 LysValSerGluValGluValGluArgAsnAlaLeuLeuAlaSerAsn------------ 1013

Qy  1862 AAGAAATGGGACTTTACAGCTTTGATCGAT------------------------------ 1891
                           |||:::|||
Db  1014 ---------------GluArgLeuMetAspAspLeuLysAsnAsnGlyGluAsnIleAla 1028

Qy  1892 TTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCT 1951
                          :::|||:::       |||::::::
Db  1029 SerLeuGlnThrGluIleGluLysLysArgAlaGluAsn--------------------- 1041

Qy  1952 GATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTG 2011
         |||    ::::::      |||   :::   :::         :::   :::   |||:::
Db  1042 AspAspLeuGlnSerLysLeuSerValValSerSerGluTyrGluAsnLeuLeuLeuIle 1061

Qy  2012 AAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCT 2071
            :::   |||   |||:::                    |||:::       :::
Db  1062 SerSerGlnThrAsnLysSerLeuGluAspLysThrAsnGlnLeuLysTyrIleGluLys 1081

Qy  2072 AATGCT------TTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTAT 2125
         |||                     :::||||||    :::::::::
Db  1082 AsnValGlnLysLeuLeuAspGluLysAspGlnArgAsnValGluLeuGluGluLeuThr 1101

Qy  2126 TCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCG 2185
         |||   |||                                  |||:::       :::
Db  1102 SerLysTyr-----------------------------GlyLysLeuGlyGluGluAsn 1111

Qy  2186 GAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGG 2245
            ::::::|||                  |||:::   ::::::|||:::
Db  1112 AlaGlnIleLysAspGluLeuLeuAlaLeuArgLysLysSerLysLysGlnHisAspLeu 1131

Qy  2246 TTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAA 2305
                  |||              |||               |||:::
Db  1132 CysAlaAsnPheValAspAspLeuLysGluLysSerAspAlaLeuGluGlnLeuThrAsn 1151

Qy  2306 CCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATG 2365
                  |||   |||:::   |||   :::             |||   :::   :::
Db  1152 GluLysAsnGluLeuIleValSerLeuGluGlnSer---AsnSerAsnAsnGluAlaLeu 1170

Qy  2366 CTACAG------GTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
         ::::::         |||   |||||||||
Db  1171 ValGluGluArgSerAspLeuAlaAsnArg----------------------------- 1180

Qy  2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
            :::   :::      :::   |||   :::   :::   :::||||||   :::
Db  1181 ---LeuSerAspMetLysLysSerLeuSerAspSerAspAsnValIleSerValIleArg 1199

Qy  2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
         ::::::::::::         |||:::   |||               ::::::   |||
Db  1200 SerAspLeuValArgValAsnAspGluLeuAspThrLeuLysLysAspLysAspSerLeu 1219

Qy  2540 TACACAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGG 2599
            |||   |||:::|||      :::                           :::|||
Db  1220 SerThrGlnTyrSerGluValCysGln--------------------------AspArg 1230

Qy  2600 GACGAC--------------------TGCGAGCGGCGC-------------------- 2617
         ||||||                    ||||||
Db  1231 AspAspLeuLeuAspSerLeuLysGlyCysGluGluSerPheAsnLysTyrAlaValSer 1250

Qy  2618 ------------TGCACT------------------------------------------ 2623
                     ||||||
Db  1251 LeuArgGluLeuCysThrLysSerGluIleAspValProValSerGluIleLeuAspAsp 1270

Qy  2624 -----------------------------GAGATCGTTGCCATCGATGCTCTTCACTTC 2653
                                      |||:::   :::            |||
Db  1271 AsnPheValPheAsnAlaGlyAsnPheSerGluLeuSerArgLeuThrValLeuSerLeu 1290

Qy  2654 AGACGCTACCTCGATCAGTTT 2674
               |||||||||   |||
Db  1291 GluAsnTyrLeuAspAlaPhe 1297
```

RESULT 14
S65236
probable membrane protein YPL217c - yeast (Saccharomyces cerevisiae)

```
N;Alternate names: hypothetical protein P1760
C;Species: Saccharomyces cerevisiae
C;Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 19-Apr-2002
C;Accession: S65236
R;Rieger, M.; Mueller-Auer, S.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65202
A;Accession: S65236
A;Molecule type: DNA
A;Residues: 1-1183 <RIE>
A;Cross-references: EMBL:Z73573; NID:g1370449; PID:e246934; PID:g1370450; GSPDB:GN00016;
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: SGD:BMS1; MIPS:YPL217c
A;Cross-references: SGD:S0006138
A;Map position: 16L
C;Keywords: transmembrane protein
F;126-142/Domain: transmembrane #status predicted <TMM>

Alignment Scores:
Pred. No.:              0.00311      Length:        1183
Score:                  159.50       Matches:       156
Percent Similarity:     31.89%       Conservative:  116
Best Local Similarity:  18.29%       Mismatches:    274
Query Match:            2.21%        Indels:        307
DB:                     2            Gaps:          36

US-09-302-812-3 (1-4069) x S65236 (1-1183)

Qy    263 AGCAGGCAGAGGCGCGTCCTCGACCCCAAGGAC--------------------------- 295
Db    341 ThrArgArgArgLysArgLeuAspAspLysAspLysLeuIleTyrAlaProMetSerAsp 360

Qy    296 ------------------------------GCTCACGTGCAGTTCAGGGTCCCACCGTCC 325
Db    361 ValGlyGlyValLeuMetAspLysAspAlaValTyrIleAspIleGlyLysLysAsnGlu 380

Qy    326 TCGCCAGCCTGCGTCCCAGGGCAG---GCGGGACAGCACAGAGGCAGCGCCACCTCGCTT 382
Db    381 GluProSerPheValProGlyGlnGluArgGlyGluGlyGluLysLeuMetThrGlyLeu 400

Qy    383 GTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATC------------ 430
Db    401 GlnSerValGluGlnSerIleAlaGluLysPheAspGlyValGlyLeuGlnLeuPheSer 420

Qy    431 ------------------AAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAAC 472
Db    421 AsnGlyThrGluLeuHisGluValAlaAspHisGluGlyMetAsp---ValGluSerGly 439

Qy    473 AATACAAGAATAGAA------------------TCCATGATGAGTTCTGTACAAAAAGAT 514
Db    440 GluGluSerIleGluAspAspGluGlyLysSerLysGlyArgThrSerLeuArgLysPro 459

Qy    515 AACTTTTACCAACATAATGTA---GAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGAT 571
Db    460 ArgIleTyrGlyLysProValGlnGluGluAspAlaAspIleAspAsnLeuProSerAsp 479

Qy    572 AAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGT 631
Db    480 GluGlu------------------ProTyrThrAsnAsp--------------------- 486

Qy    632 AAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTGGAAAGTGAACCTCAAACAGTA 691
Db    487 --------------------------AspAspValGlnAspSerGluProArgMetVal 497

Qy    692 ACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGAT 751
Db    498 GluIle---------AspPheAsnAsnThrGlyGluGlnGlyAlaGluLysLeuAlaLeu 514

Qy    752 CACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG 811
Db    515 GluThrAspSerGluPheGluGluSerGluAspGluPheSerTrpGluArgThrAla--- 533

Qy    812 CTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCC--------------------- 850
Db    534 ---AlaAsnLysLeuLysLysThrGluSerLysLysArgThrTrpAsnIleGlyLysLeu 552

Qy    851 ------AGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGAC 904
Db    553 IleTyrMetAspAsnIleSerProGluGluCysIleArgArgTrpArg---GlyGluAsp 571

Qy    905 TGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTT 964
Db    572 AspAspSer------LysAspGluSerAspIle--------------------------- 580

Qy    965 GGCTCTGAGGATGTT--------------------------GGTACTGGGTCAAAAAAT 997
Db    581 ---GluGluAspValAspAspAspPhePheArgLysLysAspGlyThrValThrLysGlu 599

Qy    998 GACAACAAA------TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAG 1051
Db    600 GlyAsnLysAspHisAlaValAspLeuGluLysPheValProTyrPheAspThrPheGlu 619

Qy   1052 AAGGAAAGTGAACCCGAATCACCGATGGATGTG--------------------------- 1084
Db    620 LysLeuAlaLysLysTrpLysSerValAspAlaIleLysGluArgPheLeuGlyAlaGly 639

Qy   1085 ---------GATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGT 1135
Db    640 IleLeuGlyAsnAspAsnLysThrLysSerAspSerAsnGluGlyGlyGluGluLeuTyr 659

Qy   1136 CCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
Db    660 GlyAspPheGluAspLeuGluAspGlyAsnProSerGluGlnAlaGluAspAsnSerAsp 679

Qy   1196 TTCCAAGCAAGAGACGCTGAC--------------------------------------- 1216
Db    680 LysGluSerGluAspGluAspGluAsnGluAspThrAsnGlyAspAspAspAsnSerPhe 699

Qy   1217 ---------------------------------------ATTGAATTTAGGAAACGGTAC 1237
Db    700 ThrAsnPheAspAlaGluGluLysLysAspLeuThrMetGluGlnGluArgGluMetAsn 719

Qy   1238 TCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACT 1297
Db    720 AlaAlaLysLysGluLysLeuArgAlaGlnPheGluIleGluGluGlyGluAsnPheLys 739

Qy   1298 GGAATGAATGATTTAAAT-----------------------------GCTAAACTACCT 1327
Db    740 GluAspAspGluAsnAsnGluTyrAspThrTrpTyrGluLeuGlnLysAlaLysIleSer 759

Qy   1328 GGAAATATTTCTAGCCTGAATGTAGAATGC---------------------AGAAATTCT 1366
Db    760 LysGlnLeuGluIleAsnAsnIleGluTyrGlnGluMetThrProGluGlnArgGlnArg 779

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCC------ 1420
Db    780 IleGluGlyPheLysAlaGlySerTyrValArgIleValPheGluLysValProMetGlu 799

Qy   1420 ------------------------------------------------------------ 1420
Db    800 PheValLysAsnPheAsnProLysPheProIleValMetGlyGlyLeuLeuProThrGlu 819

Qy   1421 ------------------AAAGCAGAGGACAGAAGAAAAGAACAGTGGGAA--------- 1453
Db    820 IleLysPheGlyIleValLysAlaArgLeuArgArg---HisArgTrpHisLysLysIle 838

Qy   1453 ------------------------------------------------------------ 1453
Db    839 LeuLysThrAsnAspProLeuValLeuSerLeuGlyTrpArgArgPheGlnThrLeuPro 858

Qy   1454 ---------ACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCAC 1504
Db    859 IleTyrThrThrThrAspSerArgThrArgThrArgMetLeuLysTyrThrProGluHis 878

Qy   1505 CTTTCTCCAGATAAGAAGTGGCTTGGA---------ACTCCCATTGAGGAGATGAGAAGA 1555
```

```
                  :::      :::   |||       :::|||
Db     879 ThrTyrCysAsnAlaAlaPheTyrGlyProLeuCysSerPro---------------Asn 893

Qy    1556 ATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT------------ 1603
              |||   ||||||:::::::::              :::    |||
Db     894 ThrProPheCysGlyValGlnIleValAlaAsnSerAspThrGlyAsnGlyPheArgIle 913

Qy    1604 ------------------CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAA 1645
                             :::    :::   ::::::::::::::        :::
Db     914 AlaAlaThrGlyIleValGluGluIleAspValAsnIleGluIleValLysLysLeuLys 933

Qy    1646 GTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAA--- 1702
           :::         ||||||        ||||||        :::          :::|||
Db     934 Leu---ValGlyPhePro------TyrLysIlePheLysAsnThrAlaPheIleLysAsp 950

Qy    1703 ATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCG 1762
           |||      |||         :::         |||            :::      :::
Db     951 MetPheSerSerAlaMetGluValAlaArgPheGluGlyAlaGlnIleLysThrValSer 970

Qy    1763 GGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAAC 1822
           |||   |||   |||   |||:::   ||||||             ::::::|||:::
Db     971 GlyIleArgGlyGlu---IleLysArgAlaLeu------------SerLysProGluGly 985

Qy    1823 ---------------TTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAA 1867
                             :::|||   ||||||      ::::::         :::
Db     986 HisTyrArgAlaAlaPheGluAspLysIleLeuMetSerAspIleValIleLeuArgSer 1005

Qy    1868 TGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAA 1927
           |||            :::      |||::::::   |||
Db    1006 TrpTyrProValArgValLysLysPheTyrAsnProValThrSer--------------- 1020

Qy    1928 CATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATT 1987

Db    1020 ------------------------------------------------------------ 1020

Qy    1988 TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAG 2047
                             |||||||||:::|||
Db    1021 ---------------LeuLeuLeuLysGluLysThrGluTrpLysGlyLeuArgLeuThr 1035

Qy    2048 GAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT 2107
              ||||||                                             |||
Db    1036 GlyGlnIle------------------------------------------ArgAla 1040

Qy    2108 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTG--------- 2158
              |||         |||      |||   ||||||      ::::::::::::
Db    1041 AlaMetAsnLeuGluThrProSerAsnProAspSerAlaTyrHisLysIleGluArgVal 1060

Qy    2159 ---------TTTGAGGGACGTTCA------------------------------------ 2173
                     |||   |||
Db    1061 GluArgHisPheAsnGlyLeuLysValProLysAlaValGlnLysGluLeuProPheLys 1080

Qy    2174 ------------TCAAGGAAACCGGAGAAACTTAAAACG 2200
                       ||||||:::|||   ||||||
Db    1081 SerGlnIleHisGlnMetLysProGlnLysLysLysThr 1093
```

RESULT 15
S38170
SRP40 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR092c; protein YKR412a
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C;Accession: S38170; S40645; S37702
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38170
A;Molecule type: DNA
A;Residues: 1-406 <BAL>
A;Cross-references: EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Ji
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40645
A;Molecule type: DNA
A;Residues: 1-406 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552
A;Experimental source: strain S288C
R;Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A;Description: Interactions between three common subunits of yeast RNA polymerases I and
A;Reference number: S37702
A;Accession: S37702
A;Molecule type: DNA
A;Residues: 1-399,'N',401-406 <LAL>
A;Cross-references: EMBL:L11275; NID:g295670; PID:g295671
C;Genetics:
A;Gene: SGD:SRP40
A;Cross-references: SGD:S0001800; MIPS:YKR092c
A;Map position: 11R

```
Alignment Scores:
Pred. No.:               0.00263      Length:          406
Score:                   159.00       Matches:         61
Percent Similarity:      40.99%       Conservative:    80
Best Local Similarity:   17.73%       Mismatches:      179
Query Match:             2.21%        Indels:          24
DB:                      2            Gaps:            4
```

US-09-302-812-3 (1-4069) x S38170 (1-406)

```
Qy     281 CTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTC 340
            |||      |||:::      ::::::   :::                ||||||     :::
Db      14 LeuSerValLysGluLysGluIleGluGluLysSerSerSerSerSerSerSerSerSer 33

Qy     341 CCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACT 400
                 :::                     |||::::::|||                :::   :::
Db      34 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerGlyGluSerSer 53

Qy     401 ATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGT 460
              :::|||         :::            :::         |||:::|||   ||||||
Db      54 SerSerSerSerSerSerSerSerSerSerSerSerAspSerSerAspSerSerAspSer 73

Qy     461 AAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTT 520
           :::   ::::::::::::               |||       ||||||         |||:::
Db      74 GluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerAspSerGlu 93

Qy     521 TACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTC 580
                       :::         :::   |||         |||            |||
Db      94 SerSerSerGluSerAspSerSerSerSerGlySerSerSerSerSerSerSerSerSer 113

Qy     581 ACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAG---TGG 637
              |||   |||::::::      ::::::   :::|||                  :::
Db     114 AspGluSerSerSerGluSerGluSerGluAspGluThrLysLysArgAlaArgGluSer 133

Qy     638 CAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAA------------ 685
              ||||||               :::      ::::::||||||:::
Db     134 AspAsnGluAspAlaLysGluThrLysLysAlaLysThrGluProGluSerSerSerSer 153

Qy     686 ------------------------ACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCT 721
                                   :::   :::            |||      |||   :::
Db     154 SerGluSerSerSerSerGlySerSerSerSerSerGluSerGluSerGlySerGluSer 173

Qy     722 AACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGA 781
           :::    |||    |||            :::         ||||||::::::|||:::
Db     174 AspSerAspSerSerSerSerSerSerSerSerSerAspSerGluSerAspSerGluSer 193

Qy     782 GACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGAT 841
           |||:::|||                              :::::::::      ::::::::
Db     194 AspSerGlnSer--------------------SerSerSerSerSerSerSerAspSer 206

Qy     842 GAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAA 901
                        ::::::   |||                  |||   |||
Db     207 SerSerAspSerAspSerSerSerSerAspSerSerSerAspSerAspSerSerSerSer 226
```

```
Qy      902 GACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGAT 961
                  :::|||          :::   |||              |||
Db      227 SerSerSerSerSerSerAspSerAspSerAspSerAspSerSerSerAspSerAspSer 246

Qy      962 GTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAA 1021
               ||||||   |||      :::   |||   :::   :::                |||
Db      247 SerGlySerSerAspSerSerSerSerSerAspSerSerSer------------AspGlu 262

Qy     1022 AGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGAT 1081
            |||          :::|||          :::   :::|||:::      |||      :::
Db      263 SerThrSerSerAspSerSerAspSerAspSerAspSerAspSerGlySerSerSerGlu 282

Qy     1082 GTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGT 1141
            ::::::      :::   :::         :::|||:::|||:::|||      :::
Db      283 LeuGluThrLysGluAlaThrAlaAspGluSerLysAlaGluGluThrProAlaSerSer 302

Qy     1142 TTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAA 1201
               :::                |||||||||   :::|||||||||   :::
Db      303 AsnGluSerThrProSerAlaSerSerSerSerSerAlaAsnLysLeuAsnIleProAla 322

Qy     1202 GCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGA 1261
                  |||         |||   ::::::::::::   ::::::             :::
Db      323 GlyThrAspGluIleLysGluGlyGlnArgLysHisPheSerArgValAspArgSerLys 342

Qy     1262 TTACATTTCCAA 1273
            ::::::|||:::
Db      343 IleAsnPheGlu 346
```

Search completed: May 26, 2004, 17:17:01
Job time : 212.062 secs

```
                   ||||||   :::   ||||||   |||       |||
Db         706 rLysGlnGluAsnLysIleLysCysLeuIleHisTyrPheGlyArgIleCysArgTrpMe 726

Qy        3601 GCCAGGTTCAAGAGCGTTGGGGTTTTTGTTTTTTTCAAGACTTGTTTTTCCTTTGCAG 3658
                |||            ||||||:::   |||   |||    :::   ||||||:::
Db         726 tPro------------ThrGlyPheValSerPheGluArgLysIleLeuProLeuGlu 741
```

RESULT 2
T21138
hypothetical protein F20C5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T21138
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19381
A;Accession: T21138
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-726 <WIL>
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1
A;Experimental source: clone F20C5
C;Genetics:
A;Gene: CESP:F20C5.1
A;Map position: 4
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

Alignment Scores:

Pred. No.:	4.82e-23	Length:	726
Score:	443.50	Matches:	184
Percent Similarity:	36.62%	Conservative:	124
Best Local Similarity:	21.88%	Mismatches:	278
Query Match:	6.55%	Indels:	255
DB:	2	Gaps:	32

US-09-302-812-5 (1-3814) x T21138 (1-726)

```
Qy         622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCA 681
               ::::::::::::|||   :::   ::::::      :::      |||:::      |||
Db          14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32

Qy         682 ACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741
               |||:::               :::   |||                  :::   :::|||
Db          33 ThrMet---------LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49

Qy         742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCA 801
               :::|||            :::   :::         ||||||      |||      |||
Db          50 GluAspProGlu------GluProLysSerArgAspValPheValSerSerGlnSerSer 67

Qy         802 GATGTTGGTGCCGAGGACATTGGAACTGGACCAAAA------------------------ 837
                |||      ||||||         |||:::
Db          68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87

Qy         838 ---------------------------------------------AATGACAACAAATTG 852
                                                                  ||||||
Db          88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107

Qy         853 ACTGGACAAGAAAGCAGCCTAGGTGAT---------TCGCCTCCATTTGAGAAAGAAAGT 903
                        |||   :::|||   :::         :::   |||   |||::::::
Db         108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127

Qy         904 GAGCCTGAGTCACCAATGGATGTA------------------------------------ 927
                        :::|||   :::||||||
Db         128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspAspGluLeuValLeu 147

Qy         928 ---GACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC 984
                  :::|||::::::               :::      :::   :::      :::
Db         148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnLeuSerGlnAspLeu 167

Qy         985 TTT---GATGAGCAAGAT------------------------------------------ 999
                |||   |||:::|||:::
Db         168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln 187

Qy        1000 ------------GATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGA 1047
                           |||         :::||||||   |||:::         :::
Db         188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207

Qy        1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107
               |||||||||
Db         208 GluAlaAspSer------------------------------------------------ 211

Qy        1108 TTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGA 1167
                  |||   ||||||::::::      |||         :::               |||
Db         212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSerSer 230

Qy        1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGG------ 1221
                  |||:::::::::      |||   :::         :::         |||
Db         231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249

Qy        1222 GATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAG--------------- 1266
                  ::::::      |||   :::      ::::::
Db         250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269

Qy        1267 ---------------------GACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGA 1305
                                       ||||||         |||   :::
Db         270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle------------ 285

Qy        1306 ACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTG 1365
                                           ||||||   |||      :::
Db         286 ------------------------------ProAsnPheProAlaSerGln--------- 292

Qy        1366 GGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTA 1425
                |||         |||:::   |||      |||:::      :::
Db         293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu--------------- 307

Qy        1426 AGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTT 1485

Db         307 ------------------------------------------------------------ 307

Qy        1486 CCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCT 1545
                           |||   ::::::::
Db         308 ------------ProGlnArgTrpArgGluPhe--------------------------- 314

Qy        1546 TGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA------------ 1593
                                             |||      |||   |||
Db         315 ------------------------------AspSerArgGlyArgArgArgAspSerTyr 324

Qy        1594 ------------------------------------------------ACTGCAGGG------ 1602
                                                                |||   |||
Db         325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344

Qy        1603 ---------------------AGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAA 1641
                                    :::   ||||||
Db         345 MetPheValGlyLeuLeuHisAsnMetTrpGluPhe------------------------ 356

Qy        1642 TTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCA--------- 1692
                                        |||   :::   |||   :::
Db         357 ------------------------AspProAspIleThrTyrLysLeuProAlaLeuGlu 368

Qy        1693 ------TATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTT 1746
                     |||   :::   :::               :::      :::|||
Db         369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388

Qy        1747 GAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCA 1806
                     |||               :::   |||||||||:::         :::|||
Db         389 ArgValAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu------ArgIleTyr 406

Qy        1807 CTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCAT 1866
                  :::   ::::::
Db         407 ArgLeuValGlyAspVal---------------------------------------Glu 413

Qy        1867 TCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACA 1926
                |||   |||:::|||   :::|||   |||:::|||:::|||      ||||||
Db         414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe------ 431
```

```
Qy       1927 TTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTC 1986
                              |||:::     |||    :::|||                  |||
Db        432 ---------------AlaArgProAspSerProPheSer------------------Phe 440

Qy       1987 AATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGC 2046
                 |||:::             |||         |||||||||||    ||||||
Db        441 CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr 460

Qy       2047 TACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTT------ACAAGA 2100
              ||||||   :::::::::         |||   |||   |||:::|||      |||:::
Db        461 TyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLys 480

Qy       2101 CAGAGTCTTGAAGATTTT---CCAGAATGG---GAAAGGTGTGAAAAGCCTCTGACACGC 2154
                    :::   |||      ||||||   ::::::       :::    |||
Db        481 MetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGlu 500

Qy       2155 TTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGAT 2214
              :::             |||      ||||||                   |||||||||
Db        501 ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp 519

Qy       2215 TTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATC 2274
              |||||||||         :::||||||||||||            |||   |||||||||||||||
Db        520 PheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIle 539

Qy       2275 AGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCAC 2334
              |||||||||:::   ||||||::::::|||         |||      |||   :::
Db        540 ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln 559

Qy       2335 AATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAA 2394
                 |||    :::   |||   |||         :::|||  ||||||||||||
Db        560 LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis 579

Qy       2395 ACTTATCGTTGGGCC---------------CGAAGCCATGAAGATGGGAGTGAA---AAG 2436
              |||   :::||||||                |||   ::::::   :::|||      :::
Db        580 ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg 599

Qy       2437 GACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGC 2496
              |||   :::   |||         |||   :::||||||||||||:::   |||:::
Db        600 AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly 619

Qy       2497 ---------------TACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAAC 2541
                             :::|||               :::   ||||||:::
Db        620 SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys 639

Qy       2542 AAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCC 2601
              ||||||      ||||||:::      |||         |||:::         :::
Db        640 LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal 658

Qy       2602 ACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAG 2661
              ||||||   |||||||||||||||||||      ||||||      :::||||||
Db        659 ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln 678

Qy       2662 ATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAG 2721
              ::::::||||||      |||:::|||   :::   :::   :::||||||:::   |||
Db        679 ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu 698

Qy       2722 TTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGA 2781
              |||            :::         :::   ::::::   :::   :::|||
Db        699 LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly 718

Qy       2782 AAA 2784
              |||
Db        719 Lys 719
```

RESULT 3
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

```
Alignment Scores:
Pred. No.:               5.33e-15      Length:         364
Score:                   328.00        Matches:        105
Percent Similarity:      41.41%        Conservative:   54
Best Local Similarity:   27.34%        Mismatches:     123
Query Match:             4.84%         Indels:         102
DB:                      2             Gaps:           11

US-09-302-812-5 (1-3814) x A84726 (1-364)

Qy       1744 CTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATT 1803
              :::   :::      |||:::      :::         ||||||   |||         :::
Db          1 MetSerLysGluGluSerSerArgTrpPheAsnGluPheLeuProAlaMetAlaCysLeu 20

Qy       1804 GCACTCTGTCTGCCAAATATTTGCACCCAG------------------------------ 1833
                 |||      |||::::::
Db         21 LeuLeuArgPheProSerLeuLeuGluSerHisTyrLeuAsnSerAspAsnLeuIleAsn 40

Qy       1834 ------------CCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCA 1881
                          :::   :::|||                     |||   :::|||
Db         41 GlyThrLysThrGlyLeuArgValLeuValProAsnLysAlaGlyIleValPheLeuSer 60

Qy       1882 CAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAAT 1941
              ||||||   |||   :::||||||:::   :::|||||||||   ||||||
Db         61 GlnGluLeuIleGlyAlaLeuLeuSerCysSerPhePheCysLeuPhePro--------- 77

Qy       1942 GCCAAGATGAAATCGGAGTATTCTAGTTAC---CCAGACATTAACTTCAATCGGTTGTTT 1998
                    :::      |||::::::   |||   |||||||||::::::||||||
Db         78 ------ValAspAspArgGlySerAsnHisLeuProIleIleAsnPheAspLysLeuPhe 95

Qy       1999 ---------------GAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTC 2043
                             ||||||:::      :::      |||:::|||   :::
Db         96 GlySerLeuIleAsnThrGlyArgAsnGluHisGlnGluAsnLysIleLysCysIleIle 115

Qy       2044 TGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAG 2103
                 ||||||:::|||::::::               |||   |||:::|||   |||:::
Db        116 HisTyrPheGlnArgLeuSerSerSerIleSerProGlyPheValSerPheGluArgLys 135

Qy       2104 ------AGTCTTGAA------------------GATTTTCCAGAATGGAAAGGTGTGAA 2139
                    |||||||||                  |||         |||   :::
Db        136 IleLeuSerLeuGluGlnAspSerSerThrLeuAspGluGlyPheTrpGlyLysSerThr 155

Qy       2140 AAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGC 2199
                 |||   :::   |||         |||   ||||||
Db        156 ValAsnLeuCysProValGluValArgThrSerGlyLeuIleGluAspGlnSerValGlu 175

Qy       2200 ATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTT 2259
                 |||:::|||||||||||||||:::   :::|||||||||               |||
Db        176 AlaLeuGluValAspPheAlaAsnLysAsnLeuGlyGlyGlyAlaLeuArgLysGlyCys 195

Qy       2260 GTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACT 2319
              ||||||||||||||||||:::||||||||||||||||||      ||||||
Db        196 ValGlnGluGluIleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeu 215

Qy       2320 GAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATAC 2379
                 ::::::      |||   :::   :::   |||   |||::::::|||   |||
Db        216 ProThrMetGluValThrGluAlaIleGluValValGlyAlaGluArgPheSerLeuTyr 235

Qy       2380 ACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGAC 2439
              ||||||         :::|||   |||:::
Db        236 ThrGly---------CysPheArgLysAlaLys--------------------------- 243
```

```
Qy       2440 GATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTAC 2499
                                    |||    |||||||||||||||||||
Db        244 -----------------ThrArgIleValAlaIleAspAlaLeuArg----------- 253

Qy       2500 CTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTC 2559

Db        253 ------------------------------------------------------------ 253

Qy       2560 CTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGT 2619
                 |||||||||
Db        254 ---HisProGlyVal--------------------------------------------- 257

Qy       2620 GGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCG 2679
                                      :::|||::::::|||:::   :::|||    ::::::
Db        258 ------------SerGlnTyrLysLeuGluSerLeuLeuSerValLeuIleLeuSerSer 273

Qy       2680 GCTGAACGTGACGTGGTTTATTTCACC---------------------TTTGGGGACTCA 2718
              :::    |||    :::      :::                        ||||||
Db        274 SerGlyArgProIleArgLeuTyrMetGlySerValSerLeuGlnGlyIleGlyAspVal 293

Qy       2719 GAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTT 2778
                 ||||||   :::|||   |||          |||            :::
Db        294 ValLeuMetValGluIleLeuSerSerSerLeuPhePheAsnGlyLeuArgPheHisArg 313

Qy       2779 GGAAAAGTGTAC 2790
                 :::|||
Db        314 SerAsnLeuTyr 317
```

RESULT 4
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Alignment Scores:

Pred. No.:	0.000461	Length:	6642
Score:	176.00	Matches:	187
Percent Similarity:	34.46%	Conservative:	120
Best Local Similarity:	20.99%	Mismatches:	372
Query Match:	2.60%	Indels:	212
DB:	2	Gaps:	42

US-09-302-812-5 (1-3814) x T29757 (1-6642)

```
Qy         88 TCTGCGCCGACTGCCTCGGAC---TCCCGGAGCTTCCCT-----------GGCAGGCAG 132
              |||:::||||||      :::    |||   |||   |||            |||   :::
Db       1438 SerSerProThrLysLysGluLysSerProSerSerProThrLysLysThrGlyAspGlu 1457

Qy        133 AGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCA 192
                 :::          |||       :::|||   :::                |||       |||
Db       1458 ValLysGluLysSerProProLysSerProThrLysLysGluLysSerProGluLysPro 1477

Qy        193 GCCTGCGTCTCGGGGCGGGCGGACCGCACAGAGGC---AACGCCACCTCGTTTGTT--- 246
                    |||                   :::       :::||||||:::    |||
Db       1478 GluAspValLysSerProValLysLysGluLysSerProAspAlaThrAsnIleValGlu 1497

Qy        247 -----------------------TTCAAACAAAAGACTATTACTACTTGGATGGATACT 282
                                        :::      |||:::||||||     |||
Db       1498 ValSerSerGluThrThrIleGluLysThrGluThrThrMetThrThrGluMetThrHis 1517

Qy        283 AAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCC 342
              :::    ::::::           |||      ||||||             ::::::|||
Db       1518 GluSerGluGluSerArgThrSerValLysLysGluLysThrProGluLysValAspGlu 1537

Qy        343 ATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAG------GTGGAAAAAATTG 396
                       |||      |||    :::       |||     |||         |||::::::
Db       1538 LysProLysSerProThrLysLysAspLysSerProGluLysSerIleThrGluGluIle 1557

Qy        397 GAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA------GAAAAGAGTTCACAG 450
                                         |||||||||          |||||||||    :::
Db       1558 --------------------------LysSerProValLysLysGluLysSerProGlu 1568

Qy        451 TATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCA 510
                 :::   ::::::       :::             |||    ::::::       |||   |||
Db       1569 LysValGluGluLysProAlaSerProThrLysLysGluLysSerProGluLysProAla 1588

Qy        511 GAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTA----------------- 552
                                  ||||||                 :::|||
Db       1589 SerProThrLysLysSerGluAsnGluValLysSerProThrLysLysGluLysSerPro 1608

Qy        553 -----------------------------CCAAAGCAGCTTAGTAATGCTAACATTGGT 582
                                            ||||||:::      |||
Db       1609 GluLysSerValValGluGluLeuLysSerProLysGluLysSerProGluLysAlaAsp 1628

Qy        583 CAGTCACCCCACACT---------------------------------------------GAT 600
                    |||     :::                                            |||
Db       1629 AspLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648

Qy        601 GACCACAGTGACACAGATCATGAAGAA------GACAGAGACAATCAGCAGTTTCTTACA 654
                    |||    |||        |||:::         ::::::         ::::::      :::
Db       1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668

Qy        655 CCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGCAGGCCAGAAGCAACTGT 714
              |||    |||          ::::::                |||     :::     :::|||
Db       1669 ProThrLysLysGluSerSerProThrLysLysThrAspAspGluValLysSerProThr 1688

Qy        715 AAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTG 774
              |||                |||      |||:::|||:::
Db       1689 LysLysGluLysSerProGlnThrValGlu---------------------------- 1698

Qy        775 GATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCA 834
                          |||     |||     |||                   |||
Db       1699 -----------GluLysProAlaSerProThrLysLysGlu----------------- 1708

Qy        835 AAAAATGACAACAAATTGACTGGACAAGAA---AGCAGCCTAGGTGATTCGCCTCCATTT 891
              |||:::          |||            :::|||         |||          :::          |||
Db       1709 LysSerProGluLysSerValValGluGluValLysSerProLysGluLysSerProGlu 1728

Qy        892 GAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAA 951
              :::      |||      :::|||:::||||||                  |||     :::
Db       1729 LysAlaGluGluLysProLysSerPro---------ThrLysLysGluLysSerProGlu 1745

Qy        952 GATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC 1011
                    |||     |||:::|||          ||||||          |||:::             |||
Db       1746 LysSerAlaAlaGluGluValLysSerProThrLysLysGluLysSerProGluLysSer 1765

Qy       1012 ---------------CAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGAT 1056
                               ::::::          |||       |||          :::     ||||||
Db       1766 AlaGluGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785

Qy       1057 GGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTT 1116
                 :::::::::                  :::   |||          ::::::
Db       1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu--------- 1802

Qy       1117 GAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCT 1176
                       |||      |||       :::                 |||       :::::::::
Db       1803 ------GluLysProAlaSerProThrLysLysGluLysThrProGluLysSerAlaAla 1820

Qy       1177 AGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACA 1236
```

```
                       |||   :::|||     :::      |||     |||     |||
Db     1821 Glu---------GluLeuLysSerProThrLysLysGluLysSerProSerSerProThr 1837

Qy     1237 GATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGT---GAAGTC 1293
                                     |||:::   |||      ||||||:::       |||
Db     1838 ------------------LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851

Qy     1294 AGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAG 1353
                  ::::::  :::           ||||||       ||||||     |||    :::
Db     1852 ProGluGluLysProLysSerProThrProLysLysSerProProGlySerProLysLys 1871

Qy     1354 AAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCAT 1413
            ||||||         :::         |||     |||   |||
Db     1872 LysLys---------SerLysSerProGluAlaGluLysProPro--------------- 1883

Qy     1414 TTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGA 1473
               |||   |||                                 :::    ||||||
Db     1884 AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAlaHis 1903

Qy     1474 GCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACAT 1533
                  ||||||            |||      |||           |||    |||
Db     1904 Phe---GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLysGlu 1922

Qy     1534 GTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1593
            :::            :::    |||    :::        :::|||:::         |||
Db     1923 Ile---------ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGluPhe 1939

Qy     1594 ACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCC 1653
                                       |||   |||    :::            |||
Db     1940 ArgCysSer---------------IleAspThrThrMetPheGlySerGlyThr------ 1952

Qy     1654 CAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCT--------------- 1698
                              :::::::::        |||   |||   |||
Db     1953 ------------ValSerValValAlaSerAsnAlaAlaGlySerValGluThrLysThr 1968

Qy     1699 --------------------------------AAGAAATGGGACTTTACA--------- 1716
                                             ||||||    :::||||||
Db     1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988

Qy     1717 ------GCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT--- 1767
                        |||      |||   |||       :::      |||   |||   |||
Db     1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008

Qy     1768 ---TTATATCAG------------------------------------------TCC 1779
                  ||||||:::                                           :::
Db     2009 ProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGlyValThr 2028

Qy     1780 ATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 1839
            |||      :::      |||    :::|||    :::||||||          |||
Db     2029 IleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn------AlaGlnAspSer 2046

Qy     1840 CCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGT 1899
               :::      :::    ::::::     |||      |||   |||     |||
Db     2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSerAlaGln 2065

Qy     1900 CTTTTAGCTAATGCTTTCTTCTGCACA---------------TTTCCCCGA--------- 1935
            |||       |||               |||                |||:::
Db     2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085

Qy     1936 ------------CGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCA--------- 1974
                              |||:::    |||:::      |||    :::|||
Db     2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105

Qy     1975 ---------GACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAA 2025
                           ||||||       ::::::    |||           ||||||
Db     2106 ValLysTrpThrIleAsn---GluLysIleValGlu------GluSerArgThrIleThr 2122

Qy     2026 AAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAGAAAAAA------CCTACA 2079
               :::||||||            |||     :::::::::       |||        |||
Db     2123 ThrIleLysThrGluAspValTyrThrLeuLysIleSerAsnAlaLysIleGluGlnThr 2142

Qy     2080 GGATTGGTGACATTTACAAGACAGAGT---------------------CTTGAAGATTTT 2118
            |||    |||        |||     |||:::                         |||
Db     2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162

Qy     2119 CCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACC 2178
                              :::   |||        :::      :::|||   :::
Db     2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182

Qy     2179 ATAGAAGGCAACGGC---CGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGA 2235
                  ||||||       |||      |||:::::::|||
Db     2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp--------------------- 2195

Qy     2236 GGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAA 2295
               |||        |||                     |||:::    :::|||:::|||   :::
Db     2196 GlyProSerProGlyThr---------------GluValSerTrpLeuLeuAsnGlyGln 2210

Qy     2296 TTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG---TGTCTTATTATC 2352
               :::                  :::|||:::||||||    :::           :::
Db     2211 ProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisVal 2230

Qy     2353 ACAGGTACTGAACAGTACAGTGAATACACAGGC 2385
               |||      |||        |||    :::|||
Db     2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241
```

RESULT 5
T43523
cut17 protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Jun-2000
C;Accession: T43523; T41649; T41700
R;Morishita, J.; Matsusaka, T.; Yanagida, M.
submitted to the EMBL Data Library, August 1999
A;Description: Fission yeast cut17 is required for chromosome segregation.
A;Reference number: Z22536
A;Accession: T43523
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-997 <MOR>
A;Cross-references: EMBL:AB031034; PIDN:BAA83415.1
R;Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22007
A;Accession: T41649
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-997 <HAR>
A;Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC962.02c
A;Experimental source: strain 972h-; cosmid c962
R;Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z22010
A;Accession: T41700
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 932-997 <WED>
A;Cross-references: EMBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10c
C;Genetics:
A;Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c
A;Map position: 3L
A;Introns: 43/3

Alignment Scores:

Pred. No.:	0.000691	Length:	997
Score:	171.00	Matches:	188
Percent Similarity:	34.44%	Conservative:	133
Best Local Similarity:	20.17%	Mismatches:	344
Query Match:	2.52%	Indels:	267
DB:	2	Gaps:	45

US-09-302-812-5 (1-3814) x T43523 (1-997)

```
Qy       46 TGGGAGCCC-----------------------------------------TGCACGAAA 63
            |||||||||                                             |||
Db      173 TrpGluProAspAspAspProTyrThrGluHisLysArgArgArgAlaAspCysValPhe 192
```

```
Qy     64 GCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCT 123
Db    193 PheThrTrpLysAspProAsnSerLeuSerProThr---------LysLeuSerPheLeu 209

Qy    124 GGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCG 183
Db    210 SerThrSerAsn------IleAspProGluAspLeuThrGluAspAsnSerIleLeuPro 227

Qy    184 TCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTT 243
Db    228 ValSerPro--------------------------------------------------- 230

Qy    244 GTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAA 303
Db    231 ------------------------ThrArgAspSerThrLysSerHisLysThrLeuAsn 242

Qy    304 ---TCAGAAAGTAAAGAAAACAAC---AATACAAGAATTGACTCCATGATG--------- 348
Db    243 PheSerProSerArgLysAsnAsnLeuAsnAlaArgProLeuThrMetSerLeuTyrThr 262

Qy    349 ---AGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAATGTT 405
Db    263 AsnThrSerGluGluLysAspSer---GlnProThrArgAlaProGlnSerProThrLys 281

Qy    406 CCTCAGCTA---------------AATCTTGATAAATCACCCACAGAAAAGAGTTCACAG 450
Db    282 ProValLeuLeuThrAlaProArgArgLysAsnLysSerPro------------------ 295

Qy    451 TATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAG---------------TGGCAGAAT 495
Db    296 ---------LysLysSerLysProAlaValPheLysProValLysProIlePheSerAsp 312

Qy    496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
Db    313 GluAspGluAspAspAspAspLeuThrAlaSerGln---------------ProPheSer 327

Qy    556 AAGCAGCTTAGTAAT------GCTAACATTGGTCAGTCACCCCACACTGATGACCACAGT 609
Db    328 LysGlyIleCysAsnAspSerMetGlnValAlaLysLysAsnPheThrGluGluIlePro 347

Qy    610 GACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCA 669
Db    348 LeuLysGluAspGluLysAspAsnGluLeuGluHisLeuValSerProAlaThrSerVal 367

Qy    670 AATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCT 729
Db    368 HisThr------------------------------------------------------ 369

Qy    730 CGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAG 789
Db    370 ------ThrValSerAspIleThrGlyHisGlnSer------------------------ 379

Qy    790 AGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAA 849
Db    380 ------ValThrAspGluSerAspGlu--------------------------------- 386

Qy    850 TTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCT 909
Db    387 ---------GlnAsnAsnCysMetSerThrProProLysIleGluIleGluSerLysIle 403

Qy    910 GAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAA 969
Db    404 GluGluGluIleSerValValSer---------------LysSerLysGluIleSerSer 418

Qy    970 GAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTT 1029
Db    419 SerValSerSerValGlyLysGluGlnAsnHisThrGluLysGlnValAlaIleGluThr 438

Qy   1030 TCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGA 1089
Db    439 ProGluGlnGlnLysValGluLysGluAspGlu--------------------------- 449

Qy   1090 AGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAA------AATAATGCTGGGACCAGT 1143
Db    450 ---------------HisLeuAsnLeuGlnGlySerPheIleGluGluSerThrLysGln 464

Qy   1144 GACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCC 1203
Db    465 ProIleSerSerLysProSerThrSerSerProAspMetThrAspAlaAlaThrGlyGly 484

Qy   1204 AAGCAGCATGGAAAA------AGGGATTCT---------------------------AAA 1230
Db    485 ArgValSerSerSerPheArgAspLysIleLeuGlnThrAsnPheSerProArgSer 504

Qy   1231 ATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAA 1290
Db    505 ThrIleAspSerPheSerAsnIleSerLysLys------ArgAsnSerGluGluAlaAsn 522

Qy   1291 GTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCA 1350
Db    523 AspGluAsnAspGluThrAsnLeuLysIlePro---------------------IlePro 535

Qy   1351 GAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATC 1410
Db    536 GluLysLysArgLysPheGlnGluValLeuGlnSerLysAsnIle--------------- 550

Qy   1411 CATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTG 1470
Db    551 ---LeuValSerSerThrGluAspSerHisGluProValLysValThrGluAspSerGln 569

Qy   1471 AGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAA 1530
Db    570 ThrAlaIleHisValSerLys---------------PheGluAspLeu---GluAsnLys 583

Qy   1531 CATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTG---GAAGATGAGAATGGT 1587
Db    584 SerMetGlu---------SerGluGlnSerLeuGlnLeuLeuSerGluSerGluAsnAsp 600

Qy   1588 GAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACA 1647
Db    601 AspLys------------------ProLeuIleAspLeuIleProLeuLeuAlaIleLys 614

Qy   1648 CGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCT--------- 1698
Db    615 ArgLysAspAsnLeuValSerGlyValLeuGluLysGlyLysSerThrSerThrSerLys 634

Qy   1699 AAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAG 1758
Db    635 ThrLysPheAsp---ThrSerIleValAspPheIleGluLysProLysThrGlu------ 651

Qy   1759 GCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCA 1818
Db    652 ------------IleSerGluValLeuProGluGluLysArgLysAlaIleCys------ 665

Qy   1819 AATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATG 1878
Db    666 ------------AspGluSerGlnThrValArgValSerIleAspArgGlyValThrLys 681

Qy   1879 TCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGG 1938
Db    682 ThrArgAsp---ValSerSerProValSerAspGlu---------------LysSerGlu 695

Qy   1939 AATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTT 1998
Db    696 AsnValAsnHisGluGluAlaAsnSerGlyHisThrValMetAsnValHisSerSerLeu 715

Qy   1999 GAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGA 2058
Db    716 AspProGlnProIleValGlnProAsnGluLeuGluSer---GlySerTyrLeuLysAsp 734

Qy   2059 GTCACAGAGAAAAAA------CCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAA 2112
Db    735 LeuProAspArgAsnValGlyAsnSerGluLysValThrPheGlnGluAspAspIleAsn 754

Qy   2113 GATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAG 2172
```

```
Db      755 Ser--------------------ProLysLeuGlnSerLysAsnAsnGlnThrValGlu 767

Qy     2173 GGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTT 2232
                :::                              ||||||       ||||||
Db      768 ---AlaValAsnThrGluThrSerAspLysLeuGlnGluLysGluAlaAsn--------- 783

Qy     2233 GGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCT 2292
                                                        |||    |||
Db      784 ------------------------------------------HisGluLeuGluAsnIle 789

Qy     2293 GAATTGATTGTTTCACGGCTGTTCACTGAGGTG--------------------------- 2325
            |||   |||       :::|||   |||||||||
Db      790 GluLysIleGluGluLysLeu---ThrGluValAspLysValSerLeuSerAspAlaPhe 808

Qy     2326 -----------CTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGT 2373
                       :::   ::::::          :::         |||||          |||
Db      809 ProAspGlnGluIleLysAsnSerArgThrSerValGlnAsnGlyThrArgSerValSer 828

Qy     2374 GAATACACAGGCTATGCTGAAACTTATCGTTGGG------CCCGAAGCCATGAAGATGGG 2427
             :::  |||          ||||||   |||                        |||
Db      829 LysAsnThrProGluLysGluThrLys-ValAspLysIleAspAsnValSerLysLysAs 848

Qy     2428 AGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487
             |||:::   :::   ||||||                     :::          |||:::
Db      848 pValGluThrSerProGlySerCysGluThrSerSerAlaPheAlaLysThrTyrAlaGl 868

Qy     2488 TTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547
               :::   |||||||||:::|||   |||         ||||||:::   |||:::::::::
Db      868 uLysGluValThrSerIleAsnLeu--Pro---SerValArgLysProLeuAspGluSer 886

Qy     2548 TAC--------------------------------------TGCGGATTCCTCCGTCCT 2568
            |||                                          ||||||   |||
Db      887 TyrTyrAspHisSerIleSerProPheAspProLeuCysGlnSerSerPheLeuAlaPro 906

Qy     2569 GGAGTTCCTTCTGAAAATCTTTCTGCAGTG 2598
                  |||   ::::::    |||:::
Db      907 GlnThrProValLysSerLysHisAlaLeu 916
```

RESULT 6
T16420
hypothetical protein F52C9.8b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16420
R;Favello, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F52C9.
A;Reference number: Z18511
A;Accession: T16420
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1165 <FAV>
A;Cross-references: EMBL:U39850; NID:g1055052; PID:g1055056; PIDN:AAA81058.1; CESP:F52C9
C;Genetics:
A;Gene: CESP:F52C9.8b
A;Introns: 21/1; 36/1; 61/3; 79/3; 107/3; 464/3; 566/1; 1010/3; 1029/3; 1106/3

```
Alignment Scores:
Pred. No.:              0.000987     Length:        1165
Score:                  169.00       Matches:       159
Percent Similarity:     31.39%       Conservative:  99
Best Local Similarity:  19.34%       Mismatches:    264
Query Match:            2.50%        Indels:        300
DB:                     2            Gaps:          36
```

US-09-302-812-5 (1-3814) x T16420 (1-1165)

```
Qy       22 CCCAGCATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAAGCGCGCTGGGGCGCCGCT 81
             |||   |||:::         ||||||       |||          |||:::      ::::::
Db      399 ProGlnMetAlaGlnGlnProGly------ProHisGlyAlaAlaGlnHisProSerSer 416

Qy       82 GGAACTTCTGCGCCG--------------------------------------------- 96
            ||||||   ||||||
Db      417 GlyThrValAlaProMetHisIleThrSerLeuProGlyAsnHisProLeuAsnArgThr 436

Qy       97 --------------------------ACTGCCTCGGACTCCCGGAGCTTC--------- 120
                                        |||   |||   |||:::
Db      437 HisLeuLeuPheAsnArgGlnValValProAlaLeuAspIleArgAsnLeuIleAlaGln 456

Qy      120 ------------------------------------------------------------ 120

Db      457 HisArgLeuMetValAspAspPheAsnProProProValHisGlnGlnProProProGln 476

Qy      121 ---CCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTC 177
               |||   :::|||:::|||                :::         |||         ||||||
Db      477 GlnProProGlnGlnGlnArgGlnLysGlnGlnArgSerGlnProAlaProAlaArgVal 496

Qy      178 CCTCCGTCCTCGCCAGCCTGCGTC---------TCGGGGCGGGCGGGACCGCACAGAGGC 228
            ||||||   |||:::   |||                 |||   |||
Db      497 ProProGlnValProSerGlnValProValThrGlyGlyValAlaAlaAspGluProPro 516

Qy      229 AACGCCACCTCGTTTGTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGA 288
                     |||:::
Db      517 ProProCysSerTyr------------------------------------------Ser 522

Qy      289 CCCAAGACAGCTGAATCAGAAAGTAAA---------------------------------- 315
            |||               ||||||||||||
Db      523 ProValAlaGlnSerSerGluSerLysIleGluProValAspValLysProArgValAla 542

Qy      316 --------------------------------------------GAAAACAACAAT 327
                                                            |||
Db      543 ProValProProGlnValProValThrProThrLysProValIleThrAsnAsnLysLys 562

Qy      328 ACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTG 387
               |||||||||         |||      |||         |||    :::|||
Db      563 LysArgIleAsp------------ValValThrLeuAspGluAspAlaProArgArgVal 578

Qy      388 GAAAAATTGGAAAATGTTCCTCAGCTAAAT--------------------CTTGATAAA 426
            :::      :::   :::|||:::::::::                          |||
Db      579 GlnValLysGlnGluIleProGluValSerSerThrSerAspAlaThrLysSerAspAla 598

Qy      427 TCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGC--- 483
            :::||||||   :::   :::      |||:::         :::   |||
Db      599 AlaProThrAlaArgGlyAlaValArgIleLysGlnGluValGluSerAspValAlaPro 618

Qy      484 -----------------------AAGTGGCAG-------------------------- 492
                                   |||::::::
Db      619 AsnThrIleLeuIleSerAlaLysLysPheGluArgMetLysAlaGluAlaGluAspLys 638

Qy      493 ------------------------------------------------AATGAAGGG 501
                                                                  |||
Db      639 GluAspMetLysLysLysIleAlaAlaLeuGlnGluAlaLeuPheAsnIleGlnGluGlu 658

Qy      502 AAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAG 561
            :::      |||:::   :::|||:::                          :::|||:::
Db      659 ArgArgValGluLysGluIleAlaAlaPheAlaThrThrAsnGlnAlaValProGlnAsn 678

Qy      562 ------CTTAGTAATGCTAACATTGGTCAG----------------------------- 585
                  |||:::      |||   |||
Db      679 GlnProAlaSerSerValGlnIleAlaGlnValSerThrSerGluSerAspAlaProGly 698

Qy      586 -------------------------------TCACCCCACACTGATGACCAC------ 606
                                          ||||||   |||   ::::::
Db      699 ThrSerGluAlaAlaAlaThrGluThrMetThrSerProLysThrLysAsnAsnValIle 718

Qy      607 ------AGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATA 660
                  ::::::   :::   |||||||||   |||                ||||||
Db      719 ValGluThrGluGlyGluGlnGluGluAspGluAspGlu------------IleProIle 734

Qy      661 AAACTTGCAAATACAAAG-------------------------------CCAACAGTA 687
            |||    :::   :::                                         |||
Db      735 LysLysSerLysLysArgArgAlaLysIleValSerAsnAspGluGluGluGluProVal 754

Qy      688 GGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGT------------GGATCTCGCCAG 735
                  :::   ||||||:::   |||                           :::
Db      755 ArgHisProLysArgArgSerAspGluLysArgGluLysArgHisValSerTyrAlaGlu 774
```

```
Qy    736 TCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCT 795
Db    775 SerAspAspAspMetProValValLysLysLysArgArgAsnGlnSerProGluAspPro 794

Qy    796 -----------------TTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAA 837
Db    795 GluTyrSerAlaAlaSerProSerGluAspGluAspAspAspIleGlySerPheValVal 814

Qy    838 AATGACAACAAATTGACTGGACAAGAAAGC---AGCCTAGGTGATTCGCCTCCATTTGAG 894
Db    815 SerAspAsnGluAspAspAspAlaAspSerPheValValGlyAspAspGluProIleGlu 834

Qy    895 AAAGAAAGTGAGCCTGAGTCACCAATG---GATGTAGACAACTCGAGAAACAGTTGTCAA 951
Db    835 TyrGluGluGluAspGluAspAspMetIleGluArgArgSerSerArgLysArgArgSer 854

Qy    952 GATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC 1011
Db    855 AspSerArgSerLysLysSerAlaThrProThr---AspArgArgArgSerArgAspThr 873

Qy   1012 CAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAA 1071
Db    874 ProThrGlySerArg------SerMetArgSerThrSerProAsnAspArgArgLysSer 891

Qy   1072 CGGTATTTGACTAAGGGAAGT---GAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAAT 1128
Db    892 ArgGluThrProProGlyAsnArgSerMetArgArgThrSerProSerAspGlyArgLys 911

Qy   1129 AATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTA 1188
Db    912 SerArgAspThrPro------ThrAlaSerSerSerMetSerSerSerThrLeuSerTyr 929

Qy   1189 GAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATG 1248
Db    930 ---CysLysLysSerLysGluThrProMetSerTyrGluGluIleGlu----------- 944

Qy   1249 AGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACA 1308
Db    945 ------GlnGlnLysLysAlaLysArgGlnArgAsnCysLysThrArgGluGluAsnArg 962

Qy   1309 GAAAGGAAG--------------------------------------------------- 1317
Db    963 GluArgLysArgLeuAlaGlnLeuGluGluLeuGluSerSerGluThrThrGlyValArg 982

Qy   1318 ---ATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCT 1374
Db    983 ArgThrLeuArgSerThrGlnAspAsnSerAspProLeuAspAlaSerLeuAlaThrThr 1002

Qy   1375 ATTGAGGAAATGAGAAAAAATGCCTCGG-------------------------------- 1401
Db   1003 IleGluGluPheArgLysThrLysLysLysAspAlaLysSerSerGluAsnArgAlaLys 1022

Qy   1401 ------------------------------------------------------------ 1401
Db   1023 GluLysGlnLysProMetAsnLysArgProThrSerSerAlaSerValAspSerAsnAsp 1042

Qy   1402 TGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTA 1461
Db   1043 AspGlyValHisIleProAlaLysArg--------------------------------- 1051

Qy   1462 GACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTT------------------------ 1497
Db   1052 ---MetAlaHisAlaSerSerValProGlyProSerArgSerLysProProMetIleGly 1070

Qy   1498 ---------------CCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATG 1542
Db   1071 AlaValLysAsnArgProAsnHis---ThrGluMetLeuAspLysArgAsnLysGlu--- 1088

Qy   1543 CCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGG 1602
Db   1089 ------SerGluGluLysArgArgLysAspArgAspGluLeu---GluArgLeuArgAsn 1105

Qy   1603 AGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTG 1662
Db   1106 LysLysHisThrThrGluGluGluLysIle------LysMetAlaArgLeuGlnAsnAla 1123

Qy   1663 ---------------AAGGATGCGATTCTGAAATACAATGTG------------------ 1689
Db   1124 LeuLysValValGlyLysAlaAlaGlyLeuLysAlaThrValLysLysGluLeuThrGly 1143

Qy   1690 GCATATTCTAAGAAATGGGACTTTACAGCTTTGGTT---GATTTCTGGGATAAGGTACTT 1746
Db   1144 SerProAlaLysLysGlnLysProAlaProLeuCysArgLysPheTrpIleSerArgLeu 1163

Qy   1747 GAAGAA 1752
Db   1164 GluGlu 1165
```

RESULT 7
I38073
nucleolar phosphoprotein p130 - human
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 24-Sep-1999
C;Accession: I38073; S52292
R;Pai, C.Y.; Chen, H.K.; Sheu, H.L.; Yeh, N.H.
J. Cell Sci. 108, 1911-1920, 1995
A;Title: Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein p
A;Reference number: I38073; MUID:95386590; PMID:7657714
A;Accession: I38073
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-699 <RES>
A;Cross-references: EMBL:Z34289; NID:g663007; PIDN:CAA84063.1; PID:g663008
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Alignment Scores:

Pred. No.:	0.00133	Length:	699
Score:	166.50	Matches:	124
Percent Similarity:	34.58%	Conservative:	70
Best Local Similarity:	22.10%	Mismatches:	228
Query Match:	2.46%	Indels:	140
DB:	2	Gaps:	22

US-09-302-812-5 (1-3814) x I38073 (1-699)

```
Qy     47 GGGAGCCCTGCACGA-----------------AAGCGCGCTGGGGCGCCGCT-GGAACT 87
Db     99 GlyProProAlaLysLysAlaAlaValProAlaLysArgValGlyLeuProProGlyLys 118

Qy     88 TCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGT--------- 138
Db    119 AlaAlaAlaLysAlaSerGluSerSerSer------SerGluGluSerArgAspAspAsp 136

Qy    139 ---GTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCC 195
Db    137 AspGluGluAspGlnLysLysGlnProValGlnLysGlyValLysProGlnAlaLysAla 156

Qy    196 TGCGTCTCGGGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTC------ 249
Db    157 ---------AlaLysAlaProProLysLysAlaLysSerSerAspSerAspSerAspSer 173

Qy    250 ----------------------------------AAACAAAAGACTATTACTACTTGG 273
Db    174 SerSerGluAspGluProProLysAsnGlnLysProLysIleThrProValThrValLys 193

Qy    274 ATGGATACTAAAGGA---CCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACA 330
Db    194 AlaGlnThrLysAlaProProLysProAlaArgAlaAlaProLysIleAlaAsnGlyLys 213

Qy    331 AGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAA 390
Db    214 AlaAlaSerSerSerSerSerSerSerSerSerSerSerSerAspAspSerGluGluGlu 233

Qy    391 AAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAA----------------------- 426
```

```
                 |||           |||:::       :::   |||
Db    234 LysAlaAlaAlaThrProLysLysThrValProLysLysGlnValValAlaLysAlaPro 253

Qy    427 ---------------TCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACT 471
                         :::||||||  |||||||||        :::   :::   :::
Db    254 ValLysAlaAlaThrThrProThrArgLysSerSer---------SerSerGluAspSer 270

Qy    472 GCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAG 531
          :::|||                :::|||   :::    ::::::   :::   ::::::
Db    271 SerSer---------------AspGluGluGluGluGlnLysLysProMetLysAsnLys 285

Qy    532 ---------------------CCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAAT 570
                               ||||||       |||    ||||||:::
Db    286 ProGlyProTyrSerTyrAlaProProProSerAlaProProProLysLys--------- 302

Qy    571 GCTAACATTGGTCAGTCACCC--------------------------------------- 591
             ::::::|||     |||
Db    303 ---SerLeuGlyThrGlnProProLysLysAlaValGluLysGlnGlnProValGluSer 321

Qy    592 ------CACACTGATGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAG 645
                :::    |||   ||||||:::       ||||||::::::
Db    322 SerGluAspSerSerAspGluSerAspSerSerSerGluGluGluLysLys--------- 338

Qy    646 TTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGA 705
                |||   |||            :::|||   |||               |||:::
Db    339 ------ProProThrLysAlaValValSerLysAlaThrThrLysProProProAlaLys 356

Qy    706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765
                                :::   :::|||       |||    :::
Db    357 Lys-----------------AlaAlaGluSerSerSerAspSerSerAspSerAspSer 370

Qy    766 GAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGA 825
           |||   |||       |||    |||                       |||
Db    371 SerGluAspAspGluAlaProSerLysPro-----------------------AlaGly 382

Qy    826 ACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCT 885
          |||      :::  ||||||                :::            ||||||
Db    383 ThrThrLysAsnSerSerAsnLys------------ProAlaValThrThrLysSerPro 398

Qy    886 CCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATG------------------------ 921
             :::     :::   |||:::   |||:::
Db    399 AlaValLysProAlaAlaAlaProLysGlnProValGlyGlyGlyGlnLysLeuLeuThr 418

Qy    922 ---------GATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAA 972
                   |||      :::|||                ||||||   ::::::
Db    419 ArgLysAlaAspSerSerSerSerGluGluGluSerSerSerSerGluGluGluLysThr 438

Qy    973 ACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCA 1032
                |||      :::        :::::::::     |||      |||
Db    439 LysLysMetValAlaThrThrLysProLysAlaThrAlaLysAlaAlaLeuSerLeuPro 458

Qy   1033 AGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGT 1092
           :::   ||||||   :::         |||                :::       |||
Db    459 AlaLysGlnAlaProGlnGlySerArgAspSerSerSerAspSerAspSerSerSerSer 478

Qy   1093 GAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAAT 1152
          |||                    |||   |||                  |||   :::
Db    479 Glu---------------------GluGluGluGluLysThrSerLysSerAlaValLys 491

Qy   1153 GCCAAGCCA------------TCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
             ||||||                 |||   :::   |||                :::
Db    492 LysLysProGlnLysValAlaGlyGlyAlaAlaProSerLysProAlaSerAlaLysLys 511

Qy   1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
             |||                |||          |||                |||
Db    512 GlyLysAlaGluSerSerAsnSerSerSerSerAspAsp---------------SerSer 526

Qy   1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
             |||:::  ||||||||         |||          :::   |||      :::
Db    527 GluGluGlu-GluGluLysLeuLysGlyLysGlySerProArgProGlnAlaProLysAl 546

Qy   1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
                 ||||||     ||||||:::             :::    |||    ||||||
```

```
Db    546 aAsnGlyThrSerAlaLeuThrAla-GlnAsnGlyLysAlaAlaLysAsnSerGluGluG 566

Qy   1381 GAAATGAG------AAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCT 1434
               |||        |||    |||:::|||                       :::
Db    566 luGluGluGluLysLysLysAlaAlaVal-----------------------ValValS 578

Qy   1435 GCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACC 1493
             ||||||   |||               :::   :::|||:::   :::   :::||||||
Db    578 erLysSerGlySerLeuLysLysArgLysGlnAsnGluAlaAlaLysGluAlaGluThr 597
```

RESULT 8
I39410
AF-4 protein, splice form 3 - human
N;Alternate names: serine/proline-rich FEL protein
C;Species: Homo sapiens (man)
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997
C;Accession: I39410
R;Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Gale, R.P.; Lanç
Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993
A;Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leul
A;Reference number: A47440; MUID:93281633; PMID:8506309
A;Accession: I39410
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1210 <NAK>
A;Cross-references: GB:L13773; NID:g306446; PID:g306447
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit
C;Genetics:
A;Gene: GDB:MLLT2; AF-4; FEL
A;Cross-references: GDB:136792; OMIM:159557
A;Map position: 4q21-4q21
C;Keywords: alternative splicing

```
Alignment Scores:
Pred. No.:                0.00162        Length:         1210
Score:                    166.00         Matches:        216
Percent Similarity:       33.43%         Conservative:   140
Best Local Similarity:    20.28%         Mismatches:     402
Query Match:              2.45%          Indels:         307
DB:                       2              Gaps:           48

US-09-302-812-5 (1-3814) x I39410 (1-1210)

Qy     15 TGCGGG---------TCCCAGCATGAGTGCGGGCCCCGGCTG---------------GGA 50
           ||||||          |||||||||          ||||||               |||
Db    151 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 170

Qy     51 GCCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTC 110
           :::   |||            ||||||                ||||||   ::
Db    171 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 190

Qy    111 CCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTT 170
           :   |||   |||   |||:::             |||            :::|||
Db    190 rAspSerAlaProGluArgGluLeuSerProLeuIleSerLeuProSerPro-------- 207

Qy    171 CAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAA 230
               |||||||||   ||||||            :::   :::      |||
Db    208 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnGl 226

Qy    231 CGCCACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACC 290
           ::::::                       :::                |||   ||
Db    226 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 246

Qy    291 CAAG------ACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCAT 344
           ||||           :::       ||||||         |||      ||||||::
Db    246 oLysAspLeuAlaValLysValHisAspLysGlu---------ThrProGlnAspSerLe 263

Qy    345 GATGAGTTCTGTGCAGAAAGATAAC------------------TTTTACCCACATAAGGT 386
           :::::::      |||        :::                    |||    |||
Db    263 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLysSe 283

Qy    387 GGAAAAAATTGGAAAATGTTCCTCAGCTAAATCTT------------------GATAAATC 428
                ::::::      |||           :::                   |||:::::
```

```
Db    283 rValAlaMetGlnGlnLysProThrAlaTyrValArgProMetAspGlyGlnAspGlnAl 303

Qy    429 ACCCACAGAAAAG-----------------AGTTCACAGTATTTGAACCAACAGCAGAC 470
Db    303 aProSerGluSerProGluLeuLysProLeuProGluAspTyr------ArgGlnGlnTh 321

Qy    471 T--------------------------------GCGAGTGTGTGCAAGTGGCAGAATGA 497
Db    321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341

Qy    498 AGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAA 557
Db    341 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluIleLeuLy 361

Qy    558 GCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGA 617
Db    361 sGluMetThrHisSer------TrpProProProLeuThrAlaIleHisThrProSerTh 379

Qy    618 TCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAA 677
Db    379 rAlaGluProSerLys------------PheProPheProThrLysAspSerGlnHisVa 395

Qy    678 GCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTC 737
Db    395 lSerSerValThrGlnAsnGlnLysGlnTyrAsp--------------------ThrSe 408

Qy    738 TGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGT---CC 794
Db    408 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 428

Qy    795 TTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGAC 854
Db    428 nLeuSerAsp------SerGluAsp-----------------SerAspSerGluGlnTh 440

Qy    855 TGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTC 914
Db    440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu-- 459

Qy    915 ACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAAC 974
Db    460 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 479

Qy    975 AAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAG 1034
Db    479 rSerSerAspSerGluSerGluSerSerSerSerAspSerGluGluAsnGluProLeuGl 499

Qy   1035 TTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGG---------------TATTT 1079
Db    499 uThrProAlaProGluProGluProProThrThrAsnLysTrpGlnLeuAspAsnTrpLe 519

Qy   1080 GACTAAGGGAAGTGAA--------------------------------------GTTAG 1100
Db    519 uThrLysValSerGlnProAlaAlaProGluGlyProArgSerThrGluProProAr 539

Qy   1101 ATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAG-- 1158
Db    539 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 559

Qy   1159 -------CCATCTGGAAACTCTTCTAGCCTT--------------------------- 1182
Db    559 rLysAspProProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 579

Qy   1183 -------AATGTAGAGTGTAGAAGTTCC------------------------------- 1203
Db    579 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 599

Qy   1204 ----------AAGCAG------------------------------------------- 1209
Db    599 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 619

Qy   1210 -------CATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTC 1262
Db    619 rSerLeuGlnGlyGluArgGluProGlyLeuLeuProTyrGlySerArgAspGlnThrSe 639

Qy   1263 AGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCC 1322
Db    639 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaAlaSerAsnGluPr 659

Qy   1323 AAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGA 1382
Db    659 oLys---------ProAlaValProProSerSerGlu---------------------- 668

Qy   1383 AATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCA 1442
Db    669 ----LysLysLysHisLysSerSerLeuProAlaProSer------------------- 680

Qy   1443 CACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAC 1502
Db    681 ---------------------------LysAlaLeuSerGlyProGluProAlaLysAs 691

Qy   1503 ACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTT 1562
Db    691 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 710

Qy   1563 GTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCA 1622
Db    711 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValValVa 730

Qy   1623 GACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATA 1682
Db    730 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----------- 746

Qy   1683 CAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGT 1742
Db    747 -------------------------------------------------------LysLe 748

Qy   1743 ACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAAT 1802
Db    748 uLeuSerProLeuArgAspThrProProProGlnSerLeu---------MetValLysIl 765

Qy   1803 TGCACTCTGTCTGCCAAATATTTGCACCCAGCCA------------------------- 1836
Db    765 eThrLeuAspLeuLeuSerArgIleProGlnProProGlyLysGlySerArgGlnArgLy 785

Qy   1836 ------------------------------------------------------------ 1836
Db    785 sAlaGluAspLysGlnProProAlaGlyLysLysHisSerSerGluLysArgSerSerAs 805

Qy   1837 -------ATACCACTCCTGAAACAGAAGATG------------------------AATCA 1865
Db    805 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 825

Qy   1866 TTCTGTCACGATG---------------TCACAGGAACAGATCGCCAGTCTTTTAGCTAA 1910
Db    825 sLysIleArgLeuGluLysGluIleLysSerGlnSerSerSerSerSerSerHisLy 845

Qy   1911 TGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTA 1970
Db    845 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 865

Qy   1971 CCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAA-- 2028
Db    865 uProProProProValSer---------------SerSerSerGlnLysProAlaLysPr 880

Qy   2029 ----CTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAG---------AAAAAACC 2075
Db    880 oAlaLeuLysArg------------SerArgArgGluAlaAspThrCysGlyGlnAspPr 896

Qy   2076 TACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGAT------TTTCCAGAATGGGA 2129
Db    896 oProLysSerAlaSerSerThrLysSerAsnHisLysAspSerSerIleProLysGlnAr 916

Qy   2130 AAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATA-------- 2181
Db    916 gArgValGluGlyLysGlySerArgSerSerSerGluHisLysGlySerSerGlyAspTh 936
```

```
Qy     2182 ------------------------------GAAGGCAACGGCCGA---GGCATGCTACA 2207
                                          ||||||   :::   |||     ||
Db      936 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 956

Qy     2208 GGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTT---GTACA 2264
               ||||    |||        :::                  |||   |||   :::::
Db      956 nValLysPheAspLysGln--------------------GlnAlaAspLeuHisMetAr 969

Qy     2265 AGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTG----------- 2313
            :|||      :::            ||||||:::       |||:::
Db      969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaPh 989

Qy     2314 -------TTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACA 2366
                                ||||||       ||||||  |||   |||   :::|||
Db      989 eLysTyrLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGluSe 1009

Qy     2367 GTACAGTGAA---------TACACAGGCTATGCTGAAACT------------------TA 2399
                |||            |||:::    |||:::||||||
Db     1009 rGlnSerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIl 1029

Qy     2400 TCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAG----------- 2448
                     :::|||        |||    :::             :::
Db     1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaVa 1049

Qy     2449 -------CGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCT 2501
                   ||||||    |||       ::::::           ||||||
Db     1049 lLeuCysMetArgCysGlnSerIle------LeuAsnMetAlaMetPheArgCysLysLy 1067

Qy     2502 CGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCT 2561
               |||      :::     |||      |||   ||||||||    :::
Db     1067 sAsp------IleAlaIleLysTyrSerArgThrLeuAsnLysHisPhe----------- 1081

Qy     2562 CCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGG 2621
                                 |||:::   |||   ||||||               |||
Db     1082 -------------------GluSerSerSerLysValAlaGlnAlaProSerProCysIl 1095

Qy     2622 TGCCTTTGGGGGT 2634
               |||     |||
Db     1095 eAlaSerThrGly 1099
```

RESULT 9
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71623
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1192 <GAR>
A;Cross-references: GB:AE001373; GB:AE001362; NID:g3845097; PIDN:AAC71813.1; PID:g384509
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0115w

Alignment Scores:

Pred. No.:	0.00189	Length:	1192
Score:	165.00	Matches:	85
Percent Similarity:	36.96%	Conservative:	61
Best Local Similarity:	21.52%	Mismatches:	143
Query Match:	2.44%	Indels:	106
DB:	2	Gaps:	18

US-09-302-812-5 (1-3814) x A71623 (1-1192)

```
Qy      235 ACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGAT---ACTAAAGGACCC 291
              :::   |||:::   |||   |||         |||   :::|||      ||||||
Db      235 SerAspPheMetAsnLysMetLysTyrGlnAsnThrGlnIleAspAsnAsnLysGlyLys 254

Qy      292 AAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCATGATGAGT 351
            ||||||        :::      |||         |||
Db      255 LysThrThrAsnThrMetGluLysAsnLysLysAsnArgAspLysLysHisSerLysLys 274

Qy      352 TCTGTGCAGAAAGATAACTTT---TACCCACATAAGGTGGAAAAATTGGAAAAT---GTT 405
                     |||   |||:::   |||         :::   :::   :::||||||   :::
Db      275 ArgLysThrLysGlnAsnTyrLysTyrLysLysGluAsnGlnAsnIleGluAsnHisIle 294

Qy      406 CCTCAGCTAAAT-----------------------CTTGATAAATCACCCACAGAA--- 438
            ||||||                             ||||||   :::      |||
Db      295 ProGlnSerLysTyrLysGlnGluArgIleGluIleLeuAspAspAsnGlyLysGluLeu 314

Qy      439 AAGAGTTCACAGTATTTGAACCAACAGCAG---------------------ACTGCGAGT 477
            ||||||   :::   :::   :::::::::                     |||   :::
Db      315 LysSerHisLysAsnIleLysGluGluLysGlyGlyIleGluLysThrAspThrThrAsn 334

Qy      478 GTGTGC---------------AAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTG 522
            :::                  |||   :::   |||   |||      ::::::
Db      335 IleAlaAspIleLysIleLysLysGluGluArgGluThrLysAspGluLysGluLysAsn 354

Qy      523 GCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGT 582
               :::                        |||   |||   :::         :::|||
Db      355 IleGlnGln-----------------LeuValLysAspValGlnLeuIleLysValGly 368

Qy      583 CAGTCACCCCAC---------------------------ACTGATGACCACAGTGACACA 615
            :::                                    |||||||||     ||||||
Db      369 GluGluThrLysAspAspGluLysGluAspLysGluGlyThrAspAspGluGluAspThr 388

Qy      616 GATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675
            |||   |||||||||   |||:::::::::                  :::      :::
Db      389 AspAspGluGluAspThrAspAspGluGluAspThrAspAspGluGluAspThrSerAsp 408

Qy      676 AAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAG 735
            :::   |||   ||||||   :::
Db      409 GluGluThrThrGlyAspGlnGlu------------------------------------ 416

Qy      736 TCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCT 795
                                    :::||||||   :::|||
Db      417 ------------------------AsnLysGluGluThrGluVal--------------- 423

Qy      796 TTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACT 855
                                                         |||   |||   |||
Db      424 ---------------------------------------------AspGluLysLysThr 428

Qy      856 ------GGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCT 909
                  :::|||      |||               ||||||:::   |||
Db      429 GluLysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGlu--- 447

Qy      910 GAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAA 969
            ||||||      |||   :::   |||            ::::::||||||   |||:::
Db      448 GluSerGluLysAspLysGluGluSerGluLysAspLysGluGluSerGluLysAspLys 467

Qy      970 GAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTT 1029
            |||   :::           |||:::|||:::         ::::::         |||
Db      468 GluLysThr-----------GluGluAspGluGluLysThrGluAsp---GluLysGly 482

Qy     1030 TCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGA 1089
            :::               :::|||   |||   |||   :::|||
Db      483 ThrGluValTyrLysLysGluThrAspValAspGluLysLysGlu-------------- 497

Qy     1090 AGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTA 1149
                                       :::||||||      ||||||   |||
Db      498 --------------------------LysGlyGluTyrGlyGluGlyThrAspAspGlu 508

Qy     1150 AATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAG 1209
                  |||                              |||   :::      ::::::
Db      509 GluAspLysGluLysGluGluAspAspGlu---------GluThrLysValGluGluLys 525

Qy     1210 CATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGAC 1269
                  ::::::|||   :::      ||||||:::               |||:::
Db      526 LysThrGluLysAspGluGluGlyThrAspTyrGluGluAspThrAspAspSerAsp--- 544
```

```
Qy       1270 AGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGG 1314
                  :::    |||:::    :::|||     :::::||||||||
Db        545 ---LysAspGluGluThrLysValGluGluLysLysThrGluArg 558
```

RESULT 10
T06029
hypothetical protein T28I19.100 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06029
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z15484
A;Accession: T06029
A;Molecule type: DNA
A;Residues: 1-532 <BEV>
A;Cross-references: EMBL:AL035709; GSPDB:GN00062; ATSP:T28I19.100
A;Experimental source: cultivar Columbia; BAC clone T28I19
C;Genetics:
A;Gene: ATSP:T28I19.100
A;Map position: 4

```
Alignment Scores:
Pred. No.:               0.0022        Length:        532
Score:                   163.00        Matches:       83
Percent Similarity:      40.11%        Conservative:  61
Best Local Similarity:   23.12%        Mismatches:    145
Query Match:             2.41%         Indels:        70
DB:                      2             Gaps:          17
```

US-09-302-812-5 (1-3814) x T06029 (1-532)

```
Qy        301 GAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAG 360
              |||   |||||||||   |||:::                :::           ||||||
Db        189 GluGlyGluSerLysSerAsnSerThrGluAsnVal---------------SerValHis 203

Qy        361 AAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTT 420
              :::|||            |||   |||   |||   ||||||                :::
Db        204 GluAspGluSerGlyPro---LysAsnGluValLeuGluGlySerValIleLysGluVal 222

Qy        421 GATAAATCACCCACAGAAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTG 480
                    :::   ||||||         |||                ||||||:::|||   |||
Db        223 SerLeuAsnThrThrGluAsnGlySerAspAspGlyGluGlnGlnGluThrLysSer--- 241

Qy        481 TGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCG 540
                           :::   :::   |||      |||:::      :::               :::
Db        242 ---------GluLeuAspSerLysThrGlyGluLysGlyPheSerAsp---------Ser 255

Qy        541 GGGACTCCGCTACCAAAG---CAGCTTAGTAATGCTAACATT------------------ 579
                          ||||||:::      ||||||   :::|||
Db        256 AsnGlyGluLeuProGluThrAsnLeuSerThrSerAsnAlaThrGluThrThrGluSer 275

Qy        580 ------------------GGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCAT 621
                                |||   |||      :::      :::      |||   :::
Db        276 SerGlySerAspGluSerGlySerSerGlyLysSerThrGlyTyrGlnGlnThrLysAsn 295

Qy        622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCA 681
              |||||||||                                                |||
Db        296 GluGluAspGlu------------------------------------------LysGlu 301

Qy        682 ACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741
                 |||         :::      |||      |||   ||||||      :::
Db        302 LysValGlnSerSerGluGluGluSerLysValLysGluSerGlyLysAsnGlu------ 319

Qy        742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCA 801
              ||||||   :::      |||   |||      :::      ||||||
Db        320 LysAspAlaSerSerSerGlnAspGluSerLysGluGluLysProGluArgLysLysLys 339

Qy        802 GATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAA 861
              :::      :::         |||   |||   :::   :::   :::            :::
Db        340 GluGluSerSerSerGln---GlyGluGlyLysGluGluGluProGluLysArgGluLys 358

Qy        862 GAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATG 921
              |||   |||                            ::::::||||||:::   |||   |||
Db        359 GluAspSerSerSer---------------GlnGluGluSerLysGluGluProGlu 373

Qy        922 GATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCA 981
              :::   :::               :::|||   |||:::            :::||||||
Db        374 AsnLysGluLysGluAlaSerSerSerGlnGluGluAsnGluIleLysGluThrGlu--- 392

Qy        982 GTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAA 1041
                       |||:::::::::   |||||||||      |||:::              :::
Db        393 ---IleLysGluLysGluGluSerSerSerGlnGluGlyAsnGluAsnLysGluThrGlu 411

Qy       1042 GCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGA 1101
                 :::   ::::::                  :::|||               ||||||   :::
Db        412 LysLysSerSerGluSer---------GlnArgLysGluAsnThrAsnSerGluLysLys 428

Qy       1102 TTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACC--------------------- 1140
                 :::         |||   |||      :::::::::   |||
Db        429 IleGlu---GlnValGluSerThrAspSerSerAsnThrGlnLysGlyAspGluGlnLys 447

Qy       1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
                 :::|||   :::   :::   |||||||||   :::|||   |||   |||      :::
Db        448 ThrAspGluSerLysArgGluSerGlyAsnAspThrSer---AsnLysGlu------Thr 464

Qy       1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
                          |||   :::|||:::      :::               |||      :::
Db        465 GluAspAspSerSerLysThrGluSerGluLysLysGluGluAsnAsnArgAsnGlyGlu 484

Qy       1261 TCAGAGGACAGAAGAAAAGAACAA------------------TGTGAAGTCAGACAT 1299
                 :::|||:::   :::   ||||||                     |||:::   |||
Db        485 ThrGluGluThrGlnAsnGluGlnGluGlnThrLysSerAlaLeuGluIleSerHis 503
```

RESULT 11
T20532
hypothetical protein F07A11.6b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T20532; T27777
R;Palmer, S.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19287
A;Accession: T20532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2722 <WIL>
A;Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b
A;Experimental source: clone F07A11
R;Gajadsty, S.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z20417
A;Accession: T27777
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2722 <WI2>
A;Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b
A;Experimental source: clone ZK20
C;Genetics:
A;Gene: CESP:F07A11.6b
A;Map position: 2
A;Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1!

```
Alignment Scores:
Pred. No.:               0.00312       Length:        2722
Score:                   163.00        Matches:       160
Percent Similarity:      33.26%        Conservative:  124
Best Local Similarity:   18.74%        Mismatches:    290
Query Match:             2.41%         Indels:        280
DB:                      2             Gaps:          36
```

US-09-302-812-5 (1-3814) x T20532 (1-2722)

```
Qy         57 CACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTCCCGGAG 116
              ||||||:::   :::         |||         :::      |||:::|||   |||   ::
Db       1410 HisGluAsnLeuIleSerProArgGlnArg-AsnArgThrThrSerSerThrSerThrAl 1429
```

```
Qy    117 CTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGT 176
          :               :::         |||         :::   |||:::
Db   1429 aThrThrSerSerLysHisGluAlaLeuSerIleProGluLysProLeuSer-------- 1446

Qy    177 CCCTCCG----------------------------------------------------- 183
           ||||||
Db   1447 -ProProValThrAlaLysSerSerValSerSerIleAspAspProSerIleArgAspGl 1466

Qy    184 ---------------------TCCTCGCCAGCCTGCGTCTCGGGGCGG----------- 210
                               |||||         :::||||||
Db   1466 uPheSerMetAsnSerAlaAlaAspSerProMetSerThrThrGlyArgProMetValLe 1486

Qy    211 ------------------------------GCGGGACCGCACAGAGGCAACGCCACCTC 239
                                          |||    :::    |||::::::||
Db   1486 uThrLysAlaAlaMetLysAlaPheAsnSerThrProProLysLysLysAsnSerSerSe 1506

Qy    240 GTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACAGC 299
          |                                             |||         ::
Db   1506 r---------------------------------------------GlyGlnHisAspSe 1511

Qy    300 TGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCA 359
          :   |||   |||      ::::::::::::            |||      |||
Db   1511 rSerSerGlySerSerSerAspSerSerSerSerAspGlySerThrSerSerAspAspSe 1531

Qy    360 GAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCT 419
             |||:::      |||      :::     |||   :::   ::::::|||   :::
Db   1531 rSerAspAspGluValProLysGlnThrGluProValThrSerIleProValValAlaSe 1551

Qy    420 TGATAAATCACCCACAGAAAAG------------------------AGTTCACAGTATTT 455
           |||          |||                                   ||||||
Db   1551 rAspAsnGlySerProGluAsnValValValGluThrProSerIleValSerGlnThrPr 1571

Qy    456 GAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACA 515
             :::   :::      ::::::   :::          :::|||         :::|||
Db   1571 oArgGluProGluProPheThrIleSerGluGlnSerSerGlu---------SerGluPr 1588

Qy    516 GCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAA 575
                   |||   |||         :::      |||:::      ::::::   ||
Db   1588 oGluAlaValProGluCysProGluAlaSerValGluProGlnMetGluThrSerGlnAs 1608

Qy    576 CATTGGT------CAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAAGAAGA 629
          |:::                     |||   |||   |||         |||   |||
Db   1608 nValGluProValSerGluGluHisGluAspSerHisGluHisGlyAspSerGluValAl 1628

Qy    630 CAGAGACAATCAGCAGTTTCTTACCTATAAAACTTGCAAATACAAAGCCAACAGTAGG 689
             ::::::||||||         |||::::::
Db   1628 aValGluSerGlnGln------GlnProLeuGlu------------------------- 1637

Qy    690 AGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTG 749

Db   1637 ------------------------------------------------------------ 1637

Qy    750 TACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGG 809
                         |||||||||    :::   |||   |||:::   :::   ||||||
Db   1638 ----------HisGlnGluGluLysGluGluLeu---GluAsnLysIleLeuAspValAl 1653

Qy    810 TGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAG 869
            ||||||                         ::::::   ::::::   |||   |||   ||
Db   1653 aAlaGlu-----------------HisHisGluGluGlnValGlnGlyAspGluAspSe 1667

Qy    870 CCTAGGTGATTCG---CCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGT 926
          |:::      |||   |||               |||   :::|||      :::
Db   1667 rValGluSerSerIleProAlaProSerAspGluProAspProValThrGlnAlaGlnGl 1687

Qy    927 AGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTT 986
             ::::::                  |||   |||   |||:::            :::||
Db   1687 uLysSerAlaHisThrLeuIleSerAspGlnGluThrAspGlnAlaValGlnSerIlePh 1707

Qy    987 TGATGAGCAAGAT---GATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGC 1043
          |||||||::::::   |||            |||      :::         |||
Db   1707 eAspGluGluGluAlaAspGluPheProGlnTyrProAspPheGlyIleSerThrAsnGl 1727

Qy   1044 AAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATT 1103
           :::|||      |||      :::      :::         |||   :::|||
Db   1727 uLysGluValSerGly------LysAspProHisAsnIleLysProThrGlu-------- 1742

Qy   1104 GCATTTCCAATTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATC 1163
                             |||   |||   :::||||||            |||||
Db   1743 ------------------ProLeuAsnAsnGlyHisThrAspLeuLeuPheSerProSe 1756

Qy   1164 TGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGA 1223
          |   ::::::   :::                        |||   ||||||
Db   1756 rSerSerAlaHisAla------------------SerGluLysGlnSer----------- 1766

Qy   1224 TTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACA 1283
                                         :::||||||||||||      :::|||
Db   1767 -------------------------------ThrLysSerGluAspAspMetGluGluAs 1776

Qy   1284 ATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCT----- 1338
             |||:::               |||         :::   :::   |||
Db   1776 pSerGluLeuValValMetGluLysGluValProMetGluGlnValIleAlaGlnGluVa 1796

Qy   1339 -AACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCC 1397
           ::::::|||   |||                :::|||:::||||||      |||:::
Db   1796 lHisValProSerGluPro------------SerProMetGluGluGluValLysLeuGl 1812

Qy   1398 TCGGTGTGGGATCCATTTGCCTTCCTTAAGACCA---------TCTGCAAGTCACACAGT 1448
             :::         |||      |||                  |||
Db   1812 uThrSerProVal------ProLysGluGluProIleLysMetGluGluSerProGluGl 1830

Qy   1449 GACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAG---------GTTCCGAAACCTTTTCC 1499
           |||         ||||||:::            |||         |||
Db   1830 nThrProThrProAspLeuIleSerAsnAsnGluSerGlnAspThrProGlyAlaValAs 1850

Qy   1500 AACACATTACAAAGATTTGTGGGAT---------------------------------AA 1526
             |||      :::   |||                                        ::
Db   1850 nAsnHisLeuHisGluAsnHisAspAlaValGlnThrProIleGlnLeuGlnProAlaSe 1870

Qy   1527 CAAACAT---GTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAA 1583
          ::::|||   |||      |||         :::   |||            :::||
Db   1870 rGlnHisGlnValAlaGlnProSerProArgProAlaValAlaProAspSerGlnGlnAs 1890

Qy   1584 TGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATT 1643
          ||||                                  |||:::::::::
Db   1890 nGly----------------------------------ProValLeuValSerGlnGl 1898

Qy   1644 CACACGACCCCAGAACTTG------AAGGATGCGATTCTGAAATACAATGTGGCATATTC 1697
           ::::::|||      :::      :::   :::   :::      |||:::      ||
Db   1898 nSerGlnProSerProMetSerSerGlnGlnSerAspMetAlaGlnAsnLeuIleLeuSe 1918

Qy   1698 TAAGAAATGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGA 1757
          |   |||            :::   |||         |||:::   :::         ||
Db   1918 rSerLys------------AspIleAsnAspLeuAlaAlaLysLeuHisLysAsnProGl 1934

Qy   1758 GGCC---------------------------------------CAACATTTATAT----- 1773
          ||||                                       |||||||||||
Db   1934 uAlaLeuAlaGlnAlaThrArgGlyAspCysSerGlyIlePheGlnHisLeuLeuLeuHi 1954

Qy   1774 ---------------CAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCC 1817
                         |||::::::   |||:::|||:::::::::
Db   1954 sAlaGlnGlyAsnGlyGlnAsnMetThrProGluMetLeuGlnLeuLysAlaAlaPhe-- 1973

Qy   1818 AAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGAT 1877
                                  :::|||:::   |||   :::         ||
Db   1974 ------------------------PheAlaGlnGlnGlnGluAsnGluAlaAsnGlnMe 1985

Qy   1878 GTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACG 1937
          |   |||
Db   1985 tMetGln----------------------------------------------------- 1987

Qy   1938 GAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTT 1997
              ||||||||||||   :::               ||||||   :::|||:::
Db   1988 ----AlaLysMetLysGlnGln---------------ThrIleAsnLysAspArgIle-- 2000

Qy   1998 TGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAG 2057
```

```
                            ::::::    |||::::::|||    ::::::                   ||
Db       2001 ----------------LysGluGlnGluArgValLysArgMetTyrGluGluAsnGluAr 2015

Qy       2058 AGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGATTT 2117
              |       |||                                          ||||||
Db       2015 gLysValGlu-----------------------------------------GluAspAr 2021

Qy       2118 TCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTAC 2177
                    |||   :::|||   |||:::             ||||||    :::   |||
Db       2021 gArgGluLysGlnArgLysGluGluGluArgGlnArgLeu-AlaAlaAlaThrAlaAlaA 2041

Qy       2178 CATAGAAGGCAACGGCCGAGGCATGC---------------------------------- 2203
                       ||||||    :::|||
Db       2041 laThrMetAlaThrGlnLysAlaAlaGluAlaLeuLysGlnLysGlnGluValProArgH 2061

Qy       2204 -----TACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACT 2258
                :::::   :::|||    ::::::            |||             |||    |
Db       2061 isGlyPheGlnHisValLeuSerMetMetThrProGlu------------AlaArgSerL 2077

Qy       2259 TGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAAT 2296
               |||||::::::             |||    :::|||
Db       2077 euTyrGluGlnPheProGlyLeuSerSerTyrIleAsn 2089
```

RESULT 12
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: E88320
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2738 <STO>
A;Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A11
C;Genetics:
A;Gene: F07A11.6
A;Map position: 2

```
Alignment Scores:
Pred. No.:               0.00312        Length:          2738
Score:                   163.00         Matches:         164
Percent Similarity:      34.63%         Conservative:    129
Best Local Similarity:   19.39%         Mismatches:      301
Query Match:             2.41%          Indels:          253
DB:                      2              Gaps:            36

US-09-302-812-5 (1-3814) x E88320 (1-2738)

Qy         52 CCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTCC 111
               |||      :::     |||        ::::::    ::::::|||    |||:::|||
Db       1416 ProArgGlnArgAsnArgThrThrSerSerThrSerThrAlaThrThrSerSerLysHis 1435

Qy        112 CGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTC 171
                 :::               ::::::    :::  |||     |||     |||       ||||||
Db       1436 GluAlaLeuSerIleProGluLysProLeu-SerProProValThrAlaLysSerSerVa 1455

Qy        172 AGGGTCCCTC---------------------------------------CGTCCTCGCCA 192
                      |||:::
Db       1455 lSerSerIleAspAspProSerIleArgAspGluPheSerMetAsnSerAlaAlaAspSe 1475

Qy        193 GCCTGCGTCTCGGGGCGGGCGGGACCGCACA--------------------------GA 225
                |||      |||      ||||||
Db       1475 rProMetSerThrThrGlyArgProMetValLeuThrLysAlaAlaMetLysAlaPheAs 1495

Qy        226 GGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATTA---------CTACTTGGATG 276
                :::|||||||||:::          |||          ::::::          :::||||||
Db       1495 nSerThrProProLysLysValSerTyrSerLeuIleIleAspCysTyrMetLeuGlyMe 1515

Qy        277 -------------GATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAA 323
                            :::    |||            :::    |||    |||         :::::
Db       1515 tMetAlaLysAsnSerSerSerGlyGlnHisAspSerSerSerGlySerSerSerAspSe 1535

Qy        324 CAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAA 383
              ::::::               |||      |||             |||:::      |||     ::
Db       1535 rSerSerSerAspGlySerThrSerSerAspAspSerSerAspAspGluValProLysGl 1555

Qy        384 GGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAG-- 441
              :    |||     :::    ::::::|||     :::          |||             |||
Db       1555 nThrGluProValThrSerIleProValValAlaSerAspAsnGlySerProGluAsnVa 1575

Qy        442 ----------------------AGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGT 479
                                      ||||||          :::    :::          :::::
Db       1575 lValValGluThrProSerIleValSerGlnThrProArgGluProGluProPheThrIl 1595

Qy        480 GTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGC 539
              :    :::          :::|||           :::|||             |||    |||
Db       1595 eSerGluGlnSerSerGlu---------SerGluProGluAlaValProGluCysProGl 1612

Qy        540 GGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGT------CAGTCACCCCA 593
                 :::          |||:::          ::::::     |||:::                ||
Db       1612 uAlaSerValGluProGlnMetGluThrSerGlnAsnValGluProValSerGluGluHi 1632

Qy        594 CACTGATGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTAC 653
              |    |||    |||          |||    |||          ::::::||||||
Db       1632 sGluAspSerHisGluHisGlyAspSerGluValAlaValGluSerGlnGln------Gl 1650

Qy        654 ACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTG 713
                |||::::::
Db       1650 nProLeuGlu-------------------------------------------------- 1653

Qy        714 TAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGT 773
                                                                 |||||||||
Db       1654 ---------------------------------------------HisGlnGluGluLy 1658

Qy        774 GGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACC 833
               :::    |||    |||:::    :::    ||||||   ||||||
Db       1658 sGluGluLeu---GluAsnLysIleLeuAspValAlaAlaGlu----------------- 1671

Qy        834 AAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCG---CCTCCATT 890
                  ::::::    ::::::    |||    |||    |||:::         |||     |||
Db       1672 -HisHisGluGluGlnValGlnGlyAspGluAspSerValGluSerSerIleProAlaPr 1691

Qy        891 TGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCA 950
                   |||    :::|||    :::                     ::::::
Db       1691 oSerAspGluProAspProValThrGlnAlaGlnGluLysSerAlaHisThrLeuIleSe 1711

Qy        951 AGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGAT---GATCGTTC 1007
                 |||    |||    |||:::                :::|||||||||::::::    |||
Db       1711 rAspGlnGluThrAspGlnAlaValGlnSerIlePheAspGluGluGluAlaAspGluPh 1731

Qy       1008 CTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAG 1067
                 |||          :::          |||                :::|||     |||     ::
Db       1731 eProGlnTyrProAspPheGlyIleSerThrAsnGluLysGluValSerGly------Ly 1749

Qy       1068 GAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAA 1127
              :        :::        |||     :::|||
Db       1749 sAspProHisAsnIleLysProThrGlu-------------------------ProLe 1760

Qy       1128 TAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGT 1187
                |||     |||     :::||||||               ||||||    ::::::    :::
Db       1760 uAsnAsnGlyHisThrAspLeuLeuPheSerProSerSerSerAlaHisAla-------- 1777

Qy       1188 AGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGATTCTAAAATTACAGATCATTTCAT 1247
                           |||    ||||||
Db       1778 ----------SerGluLysGlnSer---------------------------------- 1782

Qy       1248 GAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAAC 1307
                         :::||||||||||||          :::|||          |||:::
Db       1783 -------ThrLysSerGluAspAspMetGluGluAspSerGluLeuValValMetGluLy 1800
```

```
Qy   1308 AGAAAGGAAGATTCCAAAATACATCCCACCT------AACCTCCCTCCAGAGAAGAAGTG 1361
Db   1800 sGluValProMetGluGlnValIleAlaGlnGluValHisValProSerGluPro----- 1818

Qy   1362 GCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTC 1421
Db   1819 -------SerProMetGluGluGluValLysLeuGluThrSerProVal------ProLy 1834

Qy   1422 CTTAAGACCA---------TCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAG 1472
Db   1834 sGluGluProIleLysMetGluGluSerProGluGlnThrProThrProAspLeuIleSe 1854

Qy   1473 AGCAGGAGAG---------GTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGA 1523
Db   1854 rAsnAsnGluSerGlnAspThrProGlyAlaValAsnAsnHisLeuHisGluAsnHisAs 1874

Qy   1524 T-------------------------------AACAAACAT---GTGAAAATGCCTTG 1547
Db   1874 pAlaValGlnThrProIleGlnLeuGlnProAlaSerGlnHisGlnValAlaGlnProSe 1894

Qy   1548 TTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAG 1607
Db   1894 rProArgProAlaValAlaProAspSerGlnGlnAsnGly-------------------- 1907

Qy   1608 GTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTG----- 1662
Db   1908 ----------------ProValLeuValSerGlnGlnSerGlnProSerProMetSerSe 1922

Qy   1663 -AAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTT 1721
Db   1922 rGlnGlnSerAspMetAlaGlnAsnLeuIleLeuSerSerLys-----------AspIl 1938

Qy   1722 GGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCC-------------------- 1761
Db   1938 eAsnAspLeuAlaAlaLysLeuHisLysAsnProGluAlaLeuAlaGlnAlaThrArgGl 1958

Qy   1762 ------------------CAACATTTATAT--------------------CAGTCCAT 1781
Db   1958 yAspCysSerGlyIlePheGlnHisLeuLeuLeuHisAlaGlnGlyAsnGlyGlnAsnMe 1978

Qy   1782 TTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACC 1841
Db   1978 tThrProGluMetLeuGlnLeuLysAlaAlaPhe------------------------- 1989

Qy   1842 ACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCT 1901
Db   1990 -PheAlaGlnGlnGlnGluAsnGluAlaAsnGlnMetMetGln----------------- 2003

Qy   1902 TTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTA 1961
Db   2004 ----------------------------------------AlaLysMetLysGlnGln-- 2009

Qy   1962 TTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTGAAGGACGTTCATCAAGGAAACC 2021
Db   2010 -------------ThrIleAsnLysAspArgIle-----------------LysGluGl 2019

Qy   2022 AGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGG 2081
Db   2019 nGluArgValLysArgMetTyrGluGluAsnGluArgLysValGlu-------------- 2034

Qy   2082 ATTGGTGACATTTACAAGACAGAGTCTTGAAGATTTCCAGAATGGGAAAGGTGTGAAAA 2141
Db   2035 ---------------------------GluAspArgArgGluLysGlnArgLysGluGl 2045

Qy   2142 GCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCAT 2201
Db   2045 uGluArgGlnArgLeu-AlaAlaAlaThrAlaAlaAlaThrMetAlaThrGlnLysAlaA 2065

Qy   2202 GC-------------------------------------TACAGGTGGATTTTGCAAA 2222
Db   2065 laGluAlaLeuLysGlnLysGlnGluValProArgHisGlyPheGlnHisValLeuSerM 2085

Qy   2223 TCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTT 2282

Db   2085 etMetThrProGlu------------AlaArgSerLeuTyrGluGlnPheProGlyLeuS 2101

Qy   2283 AATCAATCCTGAAT 2296
Db   2101 erSerTyrIleAsn 2105
```

RESULT 13
A58198
serine/proline-rich FEL protein, splice form 1 - human
N;Alternate names: AF-4 protein
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 25-Oct-1996 #text_change 01-Dec-2000
C;Accession: A58198; I52572
R;Morrissey, J.J.M.
Blood 81, 1124-1131, 1993
A;Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.
A;Reference number: I52572; MUID:93184301; PMID:8443374
A;Accession: A58198
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1213 <MOR>
A;Cross-references: GB:L25050; NID:g407323
A;Accession: I52572
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'MAAQS',13-1213 <MO2>
A;Cross-references: GB:L25050; NID:g407323; PIDN:AAA36642.1; PID:g407324
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit
C;Genetics:
A;Gene: GDB:MLLT2; AF-4; FEL
A;Cross-references: GDB:136792; OMIM:159557
A;Map position: 4q21-4q21
C;Keywords: alternative splicing

Alignment Scores:

Pred. No.:	0.00309	Length:	1213
Score:	162.00	Matches:	215
Percent Similarity:	33.21%	Conservative:	137
Best Local Similarity:	20.28%	Mismatches:	401
Query Match:	2.39%	Indels:	307
DB:	2	Gaps:	48

US-09-302-812-5 (1-3814) x A58198 (1-1213)

```
Qy     15 TGCGGG---------TCCCAGCATGAGTGCGGGCCCCGGCTG--------------GGA 50
Db    158 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 177

Qy     51 GCCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTC 110
Db    178 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 197

Qy    111 CCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTT 170
Db    197 rAspSerAlaProGluArgGluLeuSerProLeuIleSerLeuProSerPro-------- 214

Qy    171 CAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAA 230
Db    215 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnGl 233

Qy    231 CGCCACCTCGTTTGTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACC 290
Db    233 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 253

Qy    291 CAAG------ACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCAT 344
Db    253 oLysAspLeuAlaValLysValHisAspLysGlu---------ThrProGlnAspSerLe 270

Qy    345 GATGAGTTCTGTGCAGAAAGATAAC------------------TTTTACCCACATAAGGT 386
Db    270 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLysSe 290
```

```
Qy    387 GGAAAAATTGGAAAATGTTCCTCAGCTAAATCTT-----------------GATAAATC 428
Db    290 rValAlaMetGlnGlnLysProThrAlaTyrValArgProMetAspGlyGlnAspGlnAl 310
Qy    429 ACCCACAGAAAAG-----------------AGTTCACAGTATTTGAACCAACAGCAGAC 470
Db    310 aProSerGluSerProGluLeuLysProLeuProGluAspTyr------ArgGlnGlnTh 328
Qy    471 T------------------------------GCGAGTGTGTGCAAGTGGCAGAATGA 497
Db    328 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 348
Qy    498 AGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAA 557
Db    348 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluIleLeuLy 368
Qy    558 GCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGA 617
Db    368 sGluMetThrHisSer------TrpProProProLeuThrAlaIleHisThrProSerTh 386
Qy    618 TCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAA 677
Db    386 rAlaGluProSerLys------------PheProPheProThrLysAspSerGlnHisVa 402
Qy    678 GCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTC 737
Db    402 lSerSerValThrGlnAsnGlnLysGlnTyrAsp--------------------ThrSe 415
Qy    738 TGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGT---CC 794
Db    415 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 435
Qy    795 TTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGAC 854
Db    435 nLeuSerAsp------SerGluAsp-----------------SerAspSerGluGlnTh 447
Qy    855 TGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTC 914
Db    447 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu-- 466
Qy    915 ACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAAC 974
Db    467 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 486
Qy    975 AAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAG 1034
Db    486 rSerSerAspSerGluSerGluSerSerSerSerAspSerGluGluAsnGluProLeuGl 506
Qy   1035 TTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGG---------------TATTT 1079
Db    506 uThrProAlaProGluProGluProProThrThrAsnLysTrpGlnLeuAspAsnTrpLe 526
Qy   1080 GACTAAGGGAAGTGAA-------------------------------------GTTAG 1100
Db    526 uThrLysValSerGlnProAlaAlaProProGluGlyProArgSerThrGluProProAr 546
Qy   1101 ATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAG-- 1158
Db    546 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 566
Qy   1159 -------CCATCTGGAAACTCTTCTAGCCTT--------------------------- 1182
Db    566 rLysAspProProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 586
Qy   1183 -------AATGTAGAGTGTAGAAGTTCC------------------------------- 1203
Db    586 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 606
Qy   1204 ----------AAGCAG------------------------------------------- 1209
Db    606 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 626
Qy   1210 -------CATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTC 1262
Db    626 rSerLeuGlnGlyGlyArgGluProGlyLeuLeuProTyrGlySerArgAspGlnThrSe 646
Qy   1263 AGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCC 1322
Db    646 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaAlaSerAsnGluPr 666
Qy   1323 AAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGA 1382
Db    666 oLys---------ProAlaValProProSerSerGlu---------------------- 675
Qy   1383 AATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCA 1442
Db    676 ----LysLysLysHisLysSerSerLeuProAlaProSer------------------- 687
Qy   1443 CACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAC 1502
Db    688 ---------------------------LysAlaLeuSerGlyProGluProAlaLysAs 698
Qy   1503 ACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTT 1562
Db    698 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 717
Qy   1563 GTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCA 1622
Db    718 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValValVa 737
Qy   1623 GACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATA 1682
Db    737 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----------- 753
Qy   1683 CAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGT 1742
Db    754 -------------------------------------------------------LysLe 755
Qy   1743 ACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAAT 1802
Db    755 uLeuSerProLeuArgAspThrProProProGlnSerLeu---------MetValLysIl 772
Qy   1803 TGCACTCTGTCTGCCAAATATTTGCACCCAGCCA------------------------- 1836
Db    772 eThrLeuAspLeuLeuSerArgIleProGlnProProGlyLysGlySerArgGlnArgLy 792
Qy   1836 ---------------------------------------------------------- 1836
Db    792 sAlaGluAspLysGlnProProAlaGlyLysLysHisSerSerGluLysArgSerSerAs 812
Qy   1837 -------ATACCACTCCTGAAACAGAAGATG-----------------------AATCA 1865
Db    812 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 832
Qy   1866 TTCTGTCACGATG---------------TCACAGGAACAGATCGCCAGTCTTTTAGCTAA 1910
Db    832 sLysIleArgLeuGluLysGluIleLysSerGlnSerSerSerSerSerSerHisLy 852
Qy   1911 TGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTA 1970
Db    852 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 872
Qy   1971 CCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAA-- 2028
Db    872 uProProProProValSer--------------SerSerSerGlnLysProAlaLysPr 887
Qy   2029 ----CTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAG---------AAAAAACC 2075
Db    887 oAlaLeuLysArg------------SerArgArgGluAlaAspThrCysGlyGlnAspPr 903
Qy   2076 TACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGAT------TTTCCAGAATGGGA 2129
Db    903 oProLysValProAlaValProArgValAsnHisLysAspSerSerIleProLysGlnAr 923
Qy   2130 AAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATA-------- 2181
```

```
Db      923 gArgValGluGlyLysGlySerArgSerSerSerAlaAspLysGlySerSerGlyAspTh 943

Qy     2182 ------------------------------GAAGGCAACGGCCGA---GGCATGCTACA 2207
                                          ||||||   :::   |||     ||
Db      943 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 963

Qy     2208 GGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTT---GTACA 2264
             ||||   |||      :::                        |||   |||    :::::
Db      963 nValLysPheAspLysGln--------------------GlnAlaAspLeuHisMetAr 976

Qy     2265 AGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTG----------- 2313
             :|||      :::                 ||||||:::     |||:::
Db      976 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaPh 996

Qy     2314 -------TTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACA 2366
                           ||||||        |||||||||  |||   |||    :::|||
Db      996 eLysTyrLeuGluAlaValLeuSerPheAsnGluCysGlyIleAlaThrGluSerGluSe 1016

Qy     2367 GTACAGTGAA---------TACACAGGCTATGCTGAAACT-----------------TA 2399
               |||              |||:::   |||:::||||||
Db     1016 rGlnSerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIl 1036

Qy     2400 TCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAG----------- 2448
                        :::|||      |||    :::               :::
Db     1036 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaVa 1056

Qy     2449 -------CGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCT 2501
                          ||||||      |||       ::::::          ||||||
Db     1056 lLeuCysMetArgCysGlnSerIle------LeuAsnMetAlaMetPheArgCysLysLy 1074

Qy     2502 CGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCT 2561
               |||       :::       |||     |||    |||||||||    :::
Db     1074 sAsp------IleAlaIleLysTyrSerArgThrLeuAsnLysHisPhe----------- 1088

Qy     2562 CCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGT 2619
                                   |||:::   |||    ||||||              |||
Db     1089 ------------------GluSerSerSerLysValAlaGlnAlaProSerProCys 1101
```

RESULT 14
F90537
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: F90537
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: F90537
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-773 <KUR>
A;Cross-references: GB:AL445566; PID:g14089619; PIDN:CAC13379.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2060
A;Genetic code: SGC3

```
Alignment Scores:
Pred. No.:                0.00304      Length:          773
Score:                    161.50       Matches:         135
Percent Similarity:       35.95%       Conservative:    80
Best Local Similarity:    22.58%       Mismatches:      241
Query Match:              2.38%        Indels:          143
DB:                       2            Gaps:            27
```

US-09-302-812-5 (1-3814) x F90537 (1-773)

```
Qy      178 CCTCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGACCGCACAGAGGCAACGCCACC 237
             |||   :::     :::||||||:::   ::::::                    |||
Db       20 ProMetAlaLeuValSerCysValThrThrLysSer------------AsnAsnGlnIle 35

Qy      238 TCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACA 297
                                 ||||||   |||      |||                  :::
Db       36 AspProSerAlaLysGlnAsnThrLysGlnThr-----------------------Ser 47

Qy      298 GCTGAATCAGAAAGTAAAGAAAACAAC---AATACAAGAATTGACTCCATGATGAGTTCT 354
               :::|||     |||||||||||    ||||||      :::|||::::::|||
Db       48 ProGlnSerAlaProLysGluAsnAsnThrAsnThrAsnArgAsnSerIleIleSerPro 67

Qy      355 GTGCAGAAAGATAAC--------------------TTTTACCCACATAAGGTGGAAAAA 393
                   |||:::                          |||   |||   |||
Db       68 GlnAsnProAspSerSerLysThrProGluThrGlnValProProThrLysProGluAsp 87

Qy      394 TTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTAT 453
                      ||||||         |||   |||      |||    :::
Db       88 GlnAsnLysGluProGln------SerProLysAspProGluIleLysAspAsnGlyGln 105

Qy      454 TTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAA 513
               |||:::        |||         |||      ::::::         |||
Db      106 LysAsnGluGlySerLysAlaProGluIleLysAspMetSerGlnLysAspGlnAlaPro 125

Qy      514 CAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCT 573
             |||:::          :::|||                ||||||:::
Db      126 GlnValProGlnLysGlnProGluAsp------------ProLysLys------------ 137

Qy      574 AACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAAGAAGACAGA 633
                              |||   |||                  :::::::::::
Db      138 ---------------ProGluThrGlnLysProProValLysSerGluAspGlnAsnLys 152

Qy      634 GACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA------AAGCCAACAGTA 687
             :::   |||            |||   |||   |||   |||          |||   |||
Db      153 GluProGlnAspProLysAlaProGluLysGlnAlaGluThrProLysAspProGlnVal 172

Qy      688 GGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGC--------------- 732
                |||            :::|||            |||   :::
Db      173 LysAsp---------MetAlaAsnLysAsnIleGlnGlyProLysValProGluLysGln 189

Qy      733 ---CAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAG 789
                |||:::   ||||||                :::::     |||      |||:::
Db      190 AlaGlnThrProLysAspProGluIleLysAsnMetAspGlnLysAspGlnAlaProGln 209

Qy      790 AGTCCTTTGTCAGATGTTGGTGCC------------------GAGGACATTGGAACTGGA 831
                |||                     |||                :::      |||
Db      210 GluProGlnLysGlnProGluAlaSerLysLysProGluThrGlnLysProProThrAsn 229

Qy      832 CCAAAAAATGACAACAAATTGACTGGACAAGAAAGC-------------------AGCCTA 873
             |||::::::   |||         |||     ||||||:::
Db      230 ProGluSerSerAsn------ThrGlnGlnGluAsnLysGlnProGluValGlnLysThr 247

Qy      874 GGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAAC 933
                :::   |||            |||         ||||||::::::|||      :::     |||
Db      248 ProGluLysProLysAlaProLysAlaGluGluProGlnAsnProGlyAsnThrGlnAsn 267

Qy      934 TCGAGAAACAGT------------------------------------TGTCAAGAT 954
             :::   |||                                         |||:::
Db      268 AsnAlaAsnAspGlnGlyGlnThrGlnGluAspSerAsnValGlnLysTrpTrpGlnAsn 287

Qy      955 TCAGAAGCAGATGAA--------------------------------------------- 969
             |||   |||   |||
Db      288 SerGlyAlaHisGluValGluLeuThrLysIleSerSerAspGlyGlnAsnLeuAsnLeu 307

Qy      970 ---------GAAACAAGTCCA--------------------GTCTTTGATGAGCAAGAT 999
                      |||      |||                        ||||||   :::   |||
Db      308 SerPheSerGluGluLeuProGlnGlyIleThrAlaLysLeuValPheSerLysThrAsp 327

Qy     1000 GATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGC 1059
             |||:::|||   ||||||       ::::::                :::      |||
Db      328 AspLysSerGlyGlnThr---LysGluIleAspPheSerThrThrLysSerLysAspGln 346

Qy     1060 GATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA 1119
             :::            |||      |||                     |||      |||
Db      347 LysIleAsnLeuGlyGluLeuArgSerGlyLysTyrThrIleSerHisIleIlePheGly 366

Qy     1120 ------------GGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGA 1167
                         |||:::         |||   |||                    |||
```

```
Db        367 IleAlaLeuPheTyrProAsnSerAsnAsnThrPheAspTyrThr---------SerAla 383

Qy       1168 AACTCTTCT--------------------------AGCCTTAATGTAGAGTGTAGAAGT 1200
               |||                                 :::|||::::::      :::
Db        384 AsnGluAspPheLeuAlaThrArgAspPheLeuArgAsnLeuAspIleAsnValLysGlu 403

Qy       1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
                 |||::::::   ::::::   |||   |||      :::             |||
Db        404 GluLysLysAsnIleGlnGlnIleSerHisIleGlyIleAsnAspPheAsnTyrSerLeu 423

Qy       1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAG------------------ACATCA 1301
                 |||            |||      :::
Db        424 IleGluLysGluSerAsnGluAsnSerLysThrPheGluPheProGlnGluIleLeuAsp 443

Qy       1302 AAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTG 1361
              ||||||:::   :::   :::|||:::         |||         |||:::   |||
Db        444 LysAsnLysPheLysGlnAlaLysValLysPheThr---PheLysSerGlnSerGluGln 462

Qy       1362 GCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTC 1421
              |||      ||||||                :::      :::   |||   |||||||||
Db        463 AlaTyrAsnSerLysArgSerValAsnLeuAspValHisIleAspAspLysPheAlaPhe 482

Qy       1422 CTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGA 1481
                 |||||||||         ::::::                   ::::::      :::
Db        483 ThrLysThrIleGluTrpThrTyrLeuSerLeuGluSerIleTyrAsnGlnPheGluLys 502

Qy       1482 GGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAAT 1541
                 :::   :::   |||   |||  |||||||||           |||:::   :::
Db        503 ArgGluGlnArgTyr---AsnGluLeu-LysAspLeuIleIleLysLysTyrPheGlu-- 520

Qy       1542 GCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGG 1601
                  :::   :::|||        |||:::|||      |||   ::::::        ||
Db        521 -------AsnArgLysAsnThrLeuProIleGluLeuProAsnTyrLysLysGlyLeuGl 538

Qy       1602 GAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTT 1661
              ||||                :::         |||   |||   |||:::      :::||
Db        538 ySer---------------GluAsnLysGluLeuAlaLysSerThrLys---IleSerLe 552

Qy       1662 GAAGGAT--------------------------GCGATTCTGAAATACAAT 1686
              |||||||                           ||||||||||||||||
Db        552 uLysAspGluIleAspAlaThrPheLysGlyTyrGlyIleLeuLysTyrAsn 569
```

RESULT 15
A47328
natural killer cell tumor-recognition protein - human
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumo
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C;Accession: A47328
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824; PMID:8421688
A;Accession: A47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1403 <AND>
A;Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252
A;Experimental source: NK killer cells from adult blood
A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)
C;Genetics:
A;Gene: GDB:NKTR
A;Cross-references: GDB:137171; OMIM:161565
A;Map position: 3p23-3p21
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>

```
Alignment Scores:
Pred. No.:               0.00406      Length:        1403
Score:                   160.50       Matches:       143
Percent Similarity:      33.47%       Conservative:  102
Best Local Similarity:   19.54%       Mismatches:    281
Query Match:             2.37%        Indels:        206
DB:                      1            Gaps:          30

US-09-302-812-5 (1-3814) x A47328 (1-1403)

Qy         85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
              |||::::::   ::::::|||   |||||||||
Db        591 ThrAlaSerLysSerSerSerHisSerArgSer--------------------------- 601

Qy        145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
                                                      |||   :::   |||
Db        602 ---------------------------------------ArgSerLysSerArgSerSer 608

Qy        205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
                 ::::::|||   ||||||   |||
Db        609 SerLysSerGly---HisArgLysArgAla------------------------------ 617

Qy        265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAAC 324
                             :::|||   |||:::         |||:::      ||||||
Db        618 ---------------SerLysSerProArgLysThrAlaSerGlnLeuSerGluAsnLys 632

Qy        325 AATACAAGAATTGACTCCATGATGAGTTCTGTG---CAGAAAGATAACTTTTACCCACAT 381
                 :::   :::   :::   :::::::::   |||   :::|||
Db        633 ProValLysThrGluProLeuArgAlaThrMetAlaGlnAsnGluAsnValValValGln 652

Qy        382 AAGGTGGAAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCC--------- 432
                 |||         ||||||:::|||   :::   |||      ||||||
Db        653 ProVal---ValAlaGluAsnIleProValIleProLeuSerAspSerProProProSer 671

Qy        433 ---------------------------------------------------------ACA 435

Db        672 ArgTrpLysProGlyGluLysProTrpLysProSerTyrGluArgIleGlnGluMetLys 691

Qy        436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
                 |||::::::         |||      |||   |||   |||:::         :::
Db        692 AlaLysThrThrHisLeuLeuProIleGlnSerThrTyrSerLeuAlaAsnIleLysGlu 711

Qy        496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
                 |||                                    :::   :::
Db        712 ThrGly------------------------------------SerSerSerSerTyrHis 719

Qy        556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACT------GATGACCACAGT 609
              |||:::      |||:::            ||||||   ::::::      |||   |||
Db        720 LysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSerAspArgSer 739

Qy        610 GACACAGATCATGAAGAAGACAGAGAC-----------------------AATCAGCAG 645
                                   |||                          ::::::
Db        740 SerGluSerSerProArgSerArgSerArgSerSerArgSerArgSerTyrSerArgSer 759

Qy        646 TTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGA 705
              :::                  ||||||::::::                  ::::::|||
Db        760 TyrThrArgSerArgSerLeuAlaSerSer---------------HisSerArgSerArg 774

Qy        706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765
              |||               |||   ||||||            |||   :::   |||   :::
Db        775 SerProSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArg 794

Qy        766 GAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGA 825
                                   |||   :::   :::   ::::::|||   |||
Db        795 --------------------SerSerSerTyrThrSerIleSerSerAspAsp---Gly 806

Qy        826 ACTGGACCAAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCT 885
                       |||         :::   :::|||:::::::::|||:::
Db        807 ArgArgAlaLysArgArgLeuArgSerSerGlyLysLysAsnSerValSerHisLysLys 826

Qy        886 ---CCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAAC 942
                          |||                |||                  |||:::
Db        827 HisSerSerSerSerGluLysThrLeuHisSerLysTyrValLysGlyArgAspArgSer 846

Qy        943 AGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT 1002
              ||||||               ::::::      :::|||      :::            :::
Db        847 SerCysValArgLysTyrSerGluSerArgSerSerLeuAspTyrSerSerAspSerGlu 866
```

```
Qy     1003 CGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGAT 1062
            :::||||||       |||   :::         |||  |||   :::|||    |||
Db      867 GlnSerSerValGlnAlaThrGln------SerAlaGlnGluLysGluLysGlnGlyGln 884

Qy     1063 CTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGA 1122
            :::   :::   :::      :::
Db      885 MetGluArgThrHisAsnLysGln------------------------------------ 892

Qy     1123 GAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTT 1182
            |||   |||   |||                                   |||
Db      893 GluLysAsnArgGlyGlu---------------------------GluLysSerLysSer 903

Qy     1183 AATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCAT 1242
                  ||||||      ||||||:::          :::::::::   ::::::||||||
Db      904 GluArgGluCysProHisSerLysLysArgThrLeuLysGluAsn---LeuSerAspHis 922

Qy     1243 TTCATGAGAATTTCCAAGTCAGAG---------------------------GACAGAAGA 1275
                        ||||||   :::                           |||    ||
Db      923 LeuArgAsnGlySerLysProLysArgLysAsnTyrAlaGlySerLysTrpAsp-SerGl 942

Qy     1276 AAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCA 1335
            |    |||:::   :::        :::||||||   |||   ||||||
Db      942 uSerAsnSerGluArg-AspValThrLysAsnSerLysAsnAspSerHis---------- 958

Qy     1336 CCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATG 1395
                     ||||||   ::::::      |||   :::        :::|||
Db      959 --------ProSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluSerGluV 976

Qy     1396 CCTCG------------------------------------------------------- 1400
               |||
Db      976 alSerGluIleHisIleLysValLysProThrThrLysSerSerThrAsnThrSerLeuP 996

Qy     1401 --------------GTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACA 1446
                             |||                     :::      |||
Db      996 roAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerThrSerAspSerGluG 1016

Qy     1447 GTGACTGTTCG------------------GGTAGACCTTCTGAGAGCAGGAGAGGTTCC- 1487
                  ||||||                  |||:::|||   :::   :::   ||||||
Db     1016 lySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHisLysHisGlySerL 1036

Qy     1488 -------------------------GAAACCTTTTCCAACACATTACAAAGATTTGTGG 1521
                                     :::                |||  |||
Db     1036 ysGluAsnLeuLysArgGluHisThrLysLysValLysGluLysLeu-LysGlyLysLys 1055

Qy     1522 GATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTAC---------CCTGTG 1572
            |||   ||||||   |||   |||      :::|||      :::         |||:::
Db     1056 AspLysLysHis---LysAlaProLysArgLysGlnAlaPheHisTrpGlnProProLeu 1074

Qy     1573 GAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTT 1632
            |||         |||                        ::::::   |||   :::
Db     1075 GluPheGlyGluGluGluGluGluGluIleAspAspLysGlnValThrGlnGluSerLys 1094

Qy     1633 CTCAACAAATTCACA---CGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTG 1689
                  |||   :::         :::   :::||||||   |||||||||
Db     1095 GluLysLysValSerGluAsnAsnGluThrIleLysAspAsnIleLeuLysThrGluLys 1114

Qy     1690 GCA------------TATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGAT 1737
            :::            |||   |||   |||         :::            |||
Db     1115 SerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLeuAsp 1134

Qy     1738 AAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTG 1797
            :::      ::::::            :::      :::   |||
Db     1135 GlnPheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrGluGlu 1154

Qy     1798 AAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAG 1857
                  :::|||   ||||||   ||||||                           |||
Db     1155 AsnValAla---CysLeuGlnAsnIle---------------------------GlnHis 1164

Qy     1858 ATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTC 1917
            :::      ||||||         |||   :::                  |||
Db     1165 ValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMetGlu 1184

Qy     1918 TTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGT--------- 1968
                                                                        
Db     1185 IleCysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLeuGly 1203

Qy     1969 ---------------TACCCAGACATTAAC 1983
                           ||||||||||||
Db     1204 AsnAlaArgLeuAspThrProAspIleAsn 1213
```

Search completed: May 26, 2004, 17:19:14
Job time : 188.361 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 26, 2004, 15:32:43 ; Search time 31.9083 Seconds
                                          (without alignments)
                                          12447.891 Million cell updates/sec

Title:          US-09-302-812-5
Perfect score:  6773
Sequence:       1 gggggactgtgtgctgcggg..........aatcatttgtcagaaaaaaa 3814

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150053_5664/app_query.fasta_1.12437
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_128_@runat_26052004_150053_5664 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      SwissProt_42:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	2.6	6632	1	UN89_CAEEL	O01761 caenorhabdi
2	172	2.5	1647	1	PQE1_CAEEL	Q10124 caenorhabdi
3	171	2.5	997	1	BIR1_SCHPO	O14064 schizosacch
4	166.5	2.5	699	1	NP14_HUMAN	Q14978 homo sapien
5	166	2.5	1210	1	AF4_HUMAN	P51825 homo sapien
6	165	2.4	2314	1	AKA6_RAT	Q9wvc7 rattus norv
7	162	2.4	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
8	161	2.4	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
9	160.5	2.4	1462	1	NKCR_HUMAN	P30414 homo sapien
10	159	2.3	687	1	DSPP_RAT	Q62598 rattus norv
11	158	2.3	5085	1	PCLO_RAT	Q9jks6 rattus norv
12	157.5	2.3	559	1	ENL_HUMAN	Q03111 homo sapien
13	157.5	2.3	934	1	DSPP_MOUSE	P97399 mus musculu
14	157.5	2.3	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
15	157.5	2.3	2211	1	FA5_BOVIN	Q28107 bos taurus
16	156	2.3	1274	1	ENAM_MOUSE	O55196 mus musculu
17	155.5	2.3	917	1	TBC2_HUMAN	Q9byx2 homo sapien
18	155	2.3	1489	1	YGP0_YEAST	P53115 saccharomyc
19	152.5	2.3	406	1	SR40_YEAST	P32583 saccharomyc
20	152.5	2.3	5120	1	PCLO_CHICK	Q9pu36 gallus gall
21	152	2.2	1983	1	TF20_MOUSE	Q9epq8 mus musculu
22	151	2.2	913	1	IF38_HUMAN	Q99613 homo sapien
23	151	2.2	1093	1	YKD5_CAEEL	Q03563 caenorhabdi
24	151	2.2	1447	1	BUD4_YEAST	P47136 saccharomyc
25	150.5	2.2	513	1	DMP1_HUMAN	Q13316 homo sapien
26	150	2.2	489	1	DMP1_RAT	P98193 rattus norv
27	150	2.2	1070	1	PVDR_PLAVS	P22290 plasmodium
28	150	2.2	1337	1	DEXT_STRDO	P39653 streptococc
29	149.5	2.2	2128	1	SPCB_MOUSE	P15508 mus musculu
30	149.5	2.2	2137	1	SPCB_HUMAN	P11277 homo sapien
31	149.5	2.2	5596	1	MDN1_HUMAN	Q9nu22 homo sapien
32	149	2.2	1453	1	NKCR_MOUSE	P30415 mus musculu
33	148	2.2	761	1	Y218_HUMAN	Q93075 homo sapien
34	148	2.2	1358	1	SIR4_YEAST	P11978 saccharomyc
35	148	2.2	1380	1	ZMS1_YEAST	P46974 saccharomyc
36	147.5	2.2	943	1	CENC_HUMAN	Q03188 homo sapien
37	147.5	2.2	1031	1	YDG9_SCHPO	Q10496 schizosacch
38	147.5	2.2	2738	1	PGCV_RAT	Q9erb4 rattus norv
39	147	2.2	823	1	UB16_HUMAN	Q9y5t5 homo sapien
40	146.5	2.2	3122	1	DPOZ_MOUSE	Q61493 mus musculu
41	146	2.2	1395	1	SP41_YEAST	P38904 saccharomyc
42	146	2.2	3924	1	ANK2_HUMAN	Q01484 homo sapien
43	145.5	2.1	1282	1	BMS1_HUMAN	Q14692 homo sapien
44	145.5	2.1	2442	1	CBP_HUMAN	Q92793 homo sapien
45	145	2.1	574	1	MIG1_CANAL	Q9y7g2 candida alb

ALIGNMENTS

```
RESULT 1
UN89_CAEEL
ID   UN89_CAEEL     STANDARD;      PRT;  6632 AA.
AC   O01761; Q17362;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN   UNC-89 OR C09D1.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=Bristol N2;
RX   MEDLINE=96180278; PubMed=8603916;
RA   Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT   "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT   assembly, encodes a giant modular protein composed of Ig and signal
RT   transduction domains.";
RL   J. Cell Biol. 132:835-848(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z., Le T.T., Wilson R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC       lattice assembly begins with positional cues laid down in the
CC       basement membrane and muscle cell membrane. UNC-89 responds to
CC       these signals, localizes, and then participates in assembling an
CC       M-line.
CC   -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC   -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 5 RCSD domains.
```

```
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33058; AAB00542.1; -.
DR   EMBL; AF003131; AAB54132.2; -.
DR   PDB; 1FHO; 20-DEC-00.
DR   WormPep; C09D1.1; CE30426.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR007850; RCSD.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 47.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF05177; RCSD; 5.
DR   Pfam; PF00621; RhoGEF; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 23.
DR   SMART; SM00325; RhoGEF; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW   3D-structure.
FT   DOMAIN       63    127       SH3.
FT   DOMAIN      152    330       DH.
FT   DOMAIN      342    498       PH.
FT   DOMAIN      547    633       IG-LIKE C2-TYPE 1.
FT   DOMAIN      648    736       IG-LIKE C2-TYPE 2.
FT   DOMAIN      748    838       IG-LIKE C2-TYPE 3.
FT   DOMAIN      946   1033       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1044   1132       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1140   1227       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1272   1315       THR-RICH.
FT   DOMAIN     1375   1475       RCSD 1.
FT   DOMAIN     1479   1585       RCSD 2.
FT   DOMAIN     1597   1695       RCSD 3.
FT   DOMAIN     1700   1799       RCSD 4.
FT   DOMAIN     1800   1860       RCSD 5.
FT   DOMAIN     1982   2067       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2071   2163       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2171   2261       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2269   2359       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2367   2455       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2463   2564       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2563   2651       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2657   2746       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2754   2858       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2887   2980       IG-LIKE C2-TYPE 16.
FT   DOMAIN     2994   3081       IG-LIKE C2-TYPE 17.
FT   DOMAIN     3087   3183       IG-LIKE C2-TYPE 18.
FT   DOMAIN     3189   3280       IG-LIKE C2-TYPE 19.
FT   DOMAIN     3286   3376       IG-LIKE C2-TYPE 20.
FT   DOMAIN     3384   3469       IG-LIKE C2-TYPE 21.
FT   DOMAIN     3482   3572       IG-LIKE C2-TYPE 22.
FT   DOMAIN     3580   3667       IG-LIKE C2-TYPE 23.
FT   DOMAIN     3686   3777       IG-LIKE C2-TYPE 24.
FT   DOMAIN     3817   3908       IG-LIKE C2-TYPE 25.
FT   DOMAIN     3920   4009       IG-LIKE C2-TYPE 26.
FT   DOMAIN     4018   4106       IG-LIKE C2-TYPE 27.
FT   DOMAIN     4109   4201       IG-LIKE C2-TYPE 28.
FT   DOMAIN     4212   4297       IG-LIKE C2-TYPE 29.
FT   DOMAIN     4302   4387       IG-LIKE C2-TYPE 30.
FT   DOMAIN     4400   4485       IG-LIKE C2-TYPE 31.
FT   DOMAIN     4489   4580       IG-LIKE C2-TYPE 32.
FT   DOMAIN     4588   4678       IG-LIKE C2-TYPE 33.
FT   DOMAIN     4681   4771       IG-LIKE C2-TYPE 34.
FT   DOMAIN     4873   4961       IG-LIKE C2-TYPE 35.
FT   DOMAIN     4965   5057       IG-LIKE C2-TYPE 36.
FT   DOMAIN     5067   5160       IG-LIKE C2-TYPE 37.
FT   DOMAIN     5171   5260       IG-LIKE C2-TYPE 38.
FT   DOMAIN     5277   5366       IG-LIKE C2-TYPE 39.
FT   DOMAIN     5383   5472       IG-LIKE C2-TYPE 40.
FT   DOMAIN     5487   5578       IG-LIKE C2-TYPE 41.
FT   DOMAIN     5595   5685       IG-LIKE C2-TYPE 42.
FT   DOMAIN     5701   5790       IG-LIKE C2-TYPE 43.
FT   DOMAIN     5815   5904       IG-LIKE C2-TYPE 44.
FT   DOMAIN     5925   6014       IG-LIKE C2-TYPE 45.
FT   DOMAIN     6038   6130       IG-LIKE C2-TYPE 46.
FT   DOMAIN     6150   6239       IG-LIKE C2-TYPE 47.
FT   DOMAIN     6275   6368       FIBRONECTIN TYPE-III.
FT   DOMAIN     6413   6502       IG-LIKE C2-TYPE 48.
FT   DOMAIN     6507   6596       IG-LIKE C2-TYPE 49.
FT   DISULFID    568    621       POTENTIAL.
FT   DISULFID   2908   2975       POTENTIAL.
FT   DISULFID   3015   3065       POTENTIAL.
FT   DISULFID   3707   3759       POTENTIAL.
FT   DISULFID   3826   3890       POTENTIAL.
FT   DISULFID   5092   5157       POTENTIAL.
FT   DISULFID   5298   5350       POTENTIAL.
FT   DISULFID   5508   5560       POTENTIAL.
FT   DISULFID   5616   5669       POTENTIAL.
FT   DISULFID   5722   5764       POTENTIAL.
FT   DISULFID   5836   5901       POTENTIAL.
FT   DISULFID   5946   5998       POTENTIAL.
FT   DISULFID   6036   6171       POTENTIAL.
FT   DISULFID   6421   6486       POTENTIAL.
FT   CONFLICT   2137   2137       A -> P (IN REF. 1).
FT   CONFLICT   2245   2247       AKA -> PKP (IN REF. 1).
FT   CONFLICT   2258   2258       A -> P (IN REF. 1).
FT   CONFLICT   2284   2284       E -> G (IN REF. 1).
FT   CONFLICT   2297   2297       M -> I (IN REF. 1).
FT   CONFLICT   3531   3531       A -> G (IN REF. 1).
FT   CONFLICT   3884   3888       DAGEY -> RRRRI (IN REF. 1).
FT   CONFLICT   3929   3929       A -> V (IN REF. 1).
FT   CONFLICT   5134   5134       A -> P (IN REF. 1).
FT   CONFLICT   5145   5145       T -> S (IN REF. 1).
FT   CONFLICT   5185   5185       G -> A (IN REF. 1).
FT   CONFLICT   5199   5199       K -> N (IN REF. 1).
FT   CONFLICT   5202   5202       L -> F (IN REF. 1).
FT   CONFLICT   5213   5213       F -> L (IN REF. 1).
FT   CONFLICT   6178   6178       A -> G (IN REF. 1).
FT   CONFLICT   6268   6268       K -> E (IN REF. 1).
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;

Alignment Scores:
Pred. No.:              0.000432       Length:         6632
Score:                  176.00         Matches:        187
Percent Similarity:     34.46%         Conservative:   120
Best Local Similarity:  20.99%         Mismatches:     372
Query Match:            2.60%          Indels:         212
DB:                     1              Gaps:           42

US-09-302-812-5 (1-3814) x UN89_CAEEL (1-6632)

Qy         88 TCTGCGCCGACTGCCTCGGAC---TCCCGGAGCTTCCCT-----------GGCAGGCAG 132
              |||:::||||||       :::   |||   |||   |||            |||   :::
Db       1438 SerSerProThrLysLysGluLysSerProSerSerProThrLysLysThrGlyAspGlu 1457

Qy        133 AGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCA 192
                 :::       |||      :::|||   :::                |||      |||
Db       1458 ValLysGluLysSerProProLysSerProThrLysLysGluLysSerProGluLysPro 1477
```

```
Qy    193 GCCTGCGTCTCGGGGCGGGCGGGACCGCACAGAGGC---AACGCCACCTCGTTTGTT--- 246
Db   1478 GluAspValLysSerProValLysLysGluLysSerProAspAlaThrAsnIleValGlu 1497

Qy    247 ------------------------TTCAAACAAAAGACTATTACTACTTGGATGGATACT 282
Db   1498 ValSerSerGluThrThrIleGluLysThrGluThrThrMetThrThrGluMetThrHis 1517

Qy    283 AAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCC 342
Db   1518 GluSerGluGluSerArgThrSerValLysLysGluLysThrProGluLysValAspGlu 1537

Qy    343 ATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAG------GTGGAAAAATTG 396
Db   1538 LysProLysSerProThrLysLysAspLysSerProGluLysSerIleThrGluGluIle 1557

Qy    397 GAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA------GAAAAGAGTTCACAG 450
Db   1558 ---------------------------LysSerProValLysLysGluLysSerProGlu 1568

Qy    451 TATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCA 510
Db   1569 LysValGluGluLysProAlaSerProThrLysLysGluLysSerProGluLysProAla 1588

Qy    511 GAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTA------------------ 552
Db   1589 SerProThrLysLysSerGluAsnGluValLysSerProThrLysLysGluLysSerPro 1608

Qy    553 ------------------------------CCAAAGCAGCTTAGTAATGCTAACATTGGT 582
Db   1609 GluLysSerValValGluGluLeuLysSerProLysGluLysSerProGluLysAlaAsp 1628

Qy    583 CAGTCACCCCACACT------------------------------------------GAT 600
Db   1629 AspLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648

Qy    601 GACCACAGTGACACAGATCATGAAGAA------GACAGAGACAATCAGCAGTTTCTTACA 654
Db   1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668

Qy    655 CCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGCAGGCCAGAAGCAACTGT 714
Db   1669 ProThrLysLysGluSerSerProThrLysLysThrAspAspGluValLysSerProThr 1688

Qy    715 AAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTG 774
Db   1689 LysLysGluLysSerProGlnThrValGlu------------------------------ 1698

Qy    775 GATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCA 834
Db   1699 ------------GluLysProAlaSerProThrLysLysGlu------------------ 1708

Qy    835 AAAAATGACAACAAATTGACTGGACAAGAA---AGCAGCCTAGGTGATTCGCCTCCATTT 891
Db   1709 LysSerProGluLysSerValValGluGluValLysSerProLysGluLysSerProGlu 1728

Qy    892 GAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAA 951
Db   1729 LysAlaGluGluLysProLysSerPro---------ThrLysLysGluLysSerProGlu 1745

Qy    952 GATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC 1011
Db   1746 LysSerAlaAlaGluGluValLysSerProThrLysLysGluLysSerProGluLysSer 1765

Qy   1012 ---------------CAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGAT 1056
Db   1766 AlaGluGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785

Qy   1057 GGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTT 1116
Db   1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu--------- 1802

Qy   1117 GAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCT 1176
Db   1803 ------GluLysProAlaSerProThrLysLysGluLysThrProGluLysSerAlaAla 1820

Qy   1177 AGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACA 1236
Db   1821 Glu---------GluLeuLysSerProThrLysLysGluLysSerProSerSerProThr 1837

Qy   1237 GATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGT---GAAGTC 1293
Db   1838 ------------------LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851

Qy   1294 AGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAG 1353
Db   1852 ProGluGluLysProLysSerProThrProLysLysSerProProGlySerProLysLys 1871

Qy   1354 AAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGATCCAT 1413
Db   1872 LysLys---------SerLysSerProGluAlaGluLysProPro--------------- 1883

Qy   1414 TTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGA 1473
Db   1884 AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAlaHis 1903

Qy   1474 GCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACAT 1533
Db   1904 Phe---GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLysGlu 1922

Qy   1534 GTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1593
Db   1923 Ile---------ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGluPhe 1939

Qy   1594 ACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCC 1653
Db   1940 ArgCysSer---------------IleAspThrThrMetPheGlySerGlyThr------ 1952

Qy   1654 CAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCT--------------- 1698
Db   1953 ------------ValSerValValAlaSerAsnAlaAlaGlySerValGluThrLysThr 1968

Qy   1699 --------------------------------AAGAAATGGGACTTTACA---------- 1716
Db   1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988

Qy   1717 ------GCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT--- 1767
Db   1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008

Qy   1768 ---TTATATCAG---------------------------------------------TCC 1779
Db   2009 ProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGlyValThr 2028

Qy   1780 ATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 1839
Db   2029 IleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn------AlaGlnAspSer 2046

Qy   1840 CCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGT 1899
Db   2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerAlaGln 2065

Qy   1900 CTTTTAGCTAATGCTTTCTTCTGCACA---------------TTTCCCCGA--------- 1935
Db   2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085

Qy   1936 ------------CGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCA--------- 1974
Db   2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105

Qy   1975 ---------GACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAA 2025
Db   2106 ValLysTrpThrIleAsn---GluLysIleValGlu------GluSerArgThrIleThr 2122

Qy   2026 AAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAGAAAAAA------CCTACA 2079
```

```
               :::|||||         |||      ::::::::     |||       |||
Db      2123 ThrIleLysThrGluAspValTyrThrLeuLysIleSerAsnAlaLysIleGluGlnThr 2142

Qy      2080 GGATTGGTGACATTTACAAGACAGAGT--------------------CTTGAAGATTTT 2118
              |||   |||    |||  |||:::                        |||
Db      2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162

Qy      2119 CCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACC 2178
                              :::  |||     :::    :::|||   :::
Db      2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182

Qy      2179 ATAGAAGGCAACGGC---CGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGA 2235
                |||||        |||      |||::::::|||
Db      2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp-------------------- 2195

Qy      2236 GGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAA 2295
              |||      |||                   |||:::   :::|||:::|||    :::
Db      2196 GlyProSerProGlyThr---------------GluValSerTrpLeuLeuAsnGlyGln 2210

Qy      2296 TTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG---TGTCTTATTATC 2352
                :::               :::|||:::||||||   :::            :::
Db      2211 ProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisVal 2230

Qy      2353 ACAGGTACTGAACAGTACAGTGAATACACAGGC 2385
              |||     |||       |||   :::|||
Db      2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241

RESULT 2
PQE1_CAEEL
ID   PQE1_CAEEL     STANDARD;      PRT;  1647 AA.
AC   Q10124; Q8MQ26; Q8MQ27; Q8MQ28; Q8MQ29; Q8MQ30; Q8MQ31;
DT   01-FEB-1996 (Rel. 33, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   PolyQ enhancer protein 1.
GN   PQE-1 OR F52C9.8/F52C9.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND C), FUNCTION, AND SUBCELLULAR
RP   LOCATION.
RX   MEDLINE=22388320; PubMed=12486229;
RA   Faber P.W., Voisine C., King D.C., Bates E.A., Hart A.C.;
RT   "Glutamine/proline-rich PQE-1 proteins protect Caenorhabditis elegans
RT   neurons from huntingtin polyglutamine neurotoxicity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17131-17136(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Favello A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Waterston R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Protects neurons from the toxic effects of expanded
CC       poly-Q disease proteins. It is not known if this is done via
CC       participation in the pathogenic mechanism underlying poly-Q-
CC       induced neurodegeneration or acting as a genetic modifier of the
CC       age of onset or progression of neurodegeneration.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC       Name=b; Synonyms=Pqe-1A;
CC         IsoId=Q10124-1; Sequence=Displayed;
CC       Name=a; Synonyms=Pqe-1C;
CC         IsoId=Q10124-2; Sequence=VSP_008069, VSP_008070;
CC       Name=c; Synonyms=Pqe-1B;
CC         IsoId=Q10124-3; Sequence=VSP_008068;
CC       Name=d;
CC         IsoId=Q10124-4; Sequence=VSP_008065;
CC         Note=No experimental confirmation available;
CC       Name=e;
CC         IsoId=Q10124-5; Sequence=VSP_008063, VSP_008064;
CC         Note=No experimental confirmation available;
CC       Name=f;
CC         IsoId=Q10124-6; Sequence=VSP_008066, VSP_008067;
CC         Note=No experimental confirmation available;
CC   -!- DOMAIN: The Gln/Pro-rich N-terminus and the Arg/Asp/Glu/Lys-rich
CC       charged domain are critical in protecting glutamatergic ASH
CC       sensory neurons from degeneration. ASH neurons expressing isoforms
CC       lacking these domains show progressive degeneration.
CC   -!- SIMILARITY: To mammalian protein GOR, yeast YGR276C and X.laevis
CC       XPMC2.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U39850; AAM45366.1; -.
DR   EMBL; U39850; AAM45367.2; -.
DR   EMBL; U39850; AAM45368.1; -.
DR   EMBL; U39850; AAM45369.1; -.
DR   EMBL; U39850; AAM45370.1; -.
DR   EMBL; U39850; AAM45371.1; -.
DR   PIR; T16421; T16421.
DR   WormPep; F52C9.8a; CE30802.
DR   WormPep; F52C9.8b; CE30998.
DR   WormPep; F52C9.8c; CE30672.
DR   WormPep; F52C9.8d; CE30673.
DR   WormPep; F52C9.8e; CE30674.
DR   WormPep; F52C9.8f; CE30675.
DR   InterPro; IPR006055; Exonuclease.
DR   Pfam; PF00929; Exonuclease; 1.
DR   SMART; SM00479; EXOIII; 1.
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN       33    568       GLN/PRO-RICH.
FT   DOMAIN      614   1356       ARG/ASP/GLU/LYS-RICH.
FT   DOMAIN     1478   1636       EXONUCLEASE.
FT   VARSPLIC      1   1074       Missing (in isoform e).
FT                                /FTId=VSP_008063.
FT   VARSPLIC   1075   1091       DAKSSENRAKEKQKPMN -> MNVSSPSTQASHFNQFQ
FT                                (in isoform e).
FT                                /FTId=VSP_008064.
FT   VARSPLIC      1   1089       Missing (in isoform d).
FT                                /FTId=VSP_008065.
FT   VARSPLIC      1   1207       Missing (in isoform f).
FT                                /FTId=VSP_008066.
FT   VARSPLIC   1208   1299       AKKQKPAPAVPKILDFSVGRTFTAIRQTAIKLVFDTFLERD
FT                                SPNAAREAQEFELSIAKQYTDGQKYRINIGHKVAALRKENT
FT                                SGILEVNKNA -> MSTYYQSAPVLVHAHAYQPPLSPPQSP
FT                                PHGYLPFYTSNPMMCSTSPLYFSPQMVQPLPPQQHYHMYPS
FT                                TSAGPGMYYTDFNPHYVQIQISQS (in isoform f).
FT                                /FTId=VSP_008067.
FT   VARSPLIC     96   1072       Missing (in isoform c).
FT                                /FTId=VSP_008068.
FT   VARSPLIC   1073   1081       KKDAKSSEN -> VRRCGGDNV (in isoform a).
FT                                /FTId=VSP_008069.
FT   VARSPLIC   1082   1647       Missing (in isoform a).
FT                                /FTId=VSP_008070.
SQ   SEQUENCE   1647 AA;  182261 MW;  A3A66BA183B3D650 CRC64;

Alignment Scores:
Pred. No.:              0.000527       Length:         1647
Score:                  172.00         Matches:        194
Percent Similarity:     31.27%         Conservative:   130
Best Local Similarity:  18.73%         Mismatches:     367
Query Match:            2.54%          Indels:         345
DB:                     1              Gaps:           43
```

```
US-09-302-812-5 (1-3814) x PQE1_CAEEL (1-1647)

Qy   121 CCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCT 180
Db   540 ProProGlnGlnGlnArgGlnLysGlnGlnArgSerGlnProAlaProAlaArgValPro 559

Qy   181 CCGTCCTCGCCAGCCTGCGTC---------TCGGGGCGGGCGGACCGCACAGAGGCAAC 231
Db   560 ProGlnValProSerGlnValProValThrGlyGlyValAlaAlaAspGluProProPro 579

Qy   232 GCCACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCC 291
Db   580 ProCysSerTyr------------------------------------------SerPro 585

Qy   292 AAGACAGCTGAATCAGAAAGTAAA------------------------------------ 315
Db   586 ValAlaGlnSerSerGluSerLysIleGluProValAspValLysProArgValAlaPro 605

Qy   316 ---------------------------------------------GAAAACAACAATACA 330
Db   606 ValProProGlnValProValThrProThrLysProValIleThrAsnAsnLysLysLys 625

Qy   331 AGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAA 390
Db   626 ArgIleAsp-----------ValValThrLeuAspGluAspAlaProArgArgValGln 641

Qy   391 AAATTGGAAAATGTTCCTCAGCTAAAT--------------------CTTGATAAATCA 429
Db   642 ValLysGlnGluIleProGluValSerSerThrSerAspAlaThrLysSerAspAlaAla 661

Qy   430 CCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGC------ 483
Db   662 ProThrAlaArgGlyAlaValArgIleLysGlnGluValGluSerAspValAlaProAsn 681

Qy   484 ---------------------AAGTGGCAG----------------------------- 492
Db   682 ThrIleLeuIleSerAlaLysLysPheGluArgMetLysAlaGluAlaGluAspLysGlu 701

Qy   493 ------------------------------------------------AATGAAGGGAAG 504
Db   702 AspMetLysLysLysIleAlaAlaLeuGlnGluAlaLeuPheAsnIleGlnGluGluArg 721

Qy   505 CATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAG--- 561
Db   722 ArgValGluLysGluIleAlaAlaPheAlaThrThrAsnGlnAlaValProGlnAsnGln 741

Qy   562 ---CTTAGTAATGCTAACATTGGTCAG-------------------------------- 585
Db   742 ProAlaSerSerValGlnIleAlaGlnValSerThrSerGluSerAspAlaProGlyThr 761

Qy   586 -----------------------------TCACCCCACACTGATGACCAC--------- 606
Db   762 SerGluAlaAlaAlaThrGluThrMetThrSerProLysThrLysAsnAsnValIleVal 781

Qy   607 ---AGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAA 663
Db   782 GluThrGluGlyGluGlnGluGluAspGluAspGlu------------IleProIleLys 797

Qy   664 CTTGCAAATACAAAG---------------------------------CCAACAGTAGGA 690
Db   798 LysSerLysLysArgArgAlaLysIleValSerAsnAspGluGluGluGluProValArg 817

Qy   691 GATGGGCAGGCCAGAAGCAACTGTAAGTGCAGT------------GGATCTCGCCAGTCT 738
Db   818 HisProLysArgArgSerAspGluLysArgGluLysArgHisValSerTyrAlaGluSer 837

Qy   739 GTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCT--- 795
Db   838 AspAspAspMetProValValLysLysLysArgArgAsnGlnSerProGluAspProGlu 857

Qy   796 ---------------TTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAAT 840
Db   858 TyrSerAlaAlaSerProSerGluAspGluAspAspAspIleGlySerPheValValSer 877

Qy   841 GACAACAAATTGACTGGACAAGAAAGC---AGCCTAGGTGATTCGCCTCCATTTGAGAAA 897
Db   878 AspAsnGluAspAspAspAlaAspSerPheValValGlyAspAspGluProIleGluTyr 897

Qy   898 GAAAGTGAGCCTGAGTCACCAATG---GATGTAGACAACTCGAGAAACAGTTGTCAAGAT 954
Db   898 GluGluGluAspGluAspAspMetIleGluArgArgSerSerArgLysArgArgSerAsp 917

Qy   955 TCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAA 1014
Db   918 SerArgSerLysLysSerAlaThrProThr---AspArgArgArgSerArgAspThrPro 936

Qy  1015 ACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGG 1074
Db   937 ThrGlySerArg------SerMetArgSerThrSerProAsnAspArgArgLysSerArg 954

Qy  1075 TATTTGACTAAGGGAAGT---GAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAAT 1131
Db   955 GluThrProProGlyAsnArgSerMetArgArgThrSerProSerAspGlyArgLysSer 974

Qy  1132 GCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAG 1191
Db   975 ArgAspThrPro------ThrAlaSerSerSerMetSerSerSerThrLeuSerTyr--- 991

Qy  1192 TGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGA 1251
Db   992 CysLysLysSerLysGluThrProMetSerTyrGluGluIleGlu--------------- 1006

Qy  1252 ATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAA 1311
Db  1007 ---GlnGlnLysLysAlaLysArgGlnArgAsnCysLysThrArgGluGluAsnArgGlu 1025

Qy  1312 AGGAAG------------------------------------------------------ 1317
Db  1026 ArgLysArgLeuAlaGlnLeuGluGluLeuGluSerSerGluThrThrGlyValArgArg 1045

Qy  1318 ATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATT 1377
Db  1046 ThrLeuArgSerThrGlnAspAsnSerAspProLeuAspAlaSerLeuAlaThrThrIle 1065

Qy  1378 GAGGAAATGAGAAAAATGCCTCGG------------------------------------ 1401
Db  1066 GluGluPheArgLysThrLysLysLysAspAlaLysSerSerGluAsnArgAlaLysGlu 1085

Qy  1402 ---------------------------------------------------------TGT 1404
Db  1086 LysGlnLysProMetAsnLysArgProThrSerSerAlaSerValAspSerAsnAspAsp 1105

Qy  1405 GGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGAC 1464
Db  1106 GlyValHisIleProAlaLysArg------------------------------------ 1113

Qy  1465 CTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTT--------------------------- 1497
Db  1114 MetAlaHisAlaSerSerValProGlyProSerArgSerLysProProMetIleGlyAla 1133

Qy  1498 ------------CCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCT 1545
Db  1134 ValLysAsnArgProAsnHis---ThrGluMetLeuAspLysArgAsnLysGlu------ 1150

Qy  1546 TGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGT 1605
Db  1151 ---SerGluGluLysArgArgLysAspArgAspGluLeu---GluArgLeuArgAsnLys 1168

Qy  1606 AGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTG--- 1662
Db  1169 LysHisThrThrGluGluGluLysIle------LysMetAlaArgLeuGlnAsnAlaLeu 1186

Qy  1663 ------------AAGGATGCGATTCTGAAATACAATGTG--------------------- 1689
Db  1187 LysValValGlyLysAlaAlaGlyLeuLysAlaThrValLysLysGluLeuThrGlySer 1206
```

```
Qy     1690 --------------------------GCATATTCTAAGAAATGGGACTTT--------- 1713
                                      |||      |||     ||||||
Db     1207 ProAlaLysLysGlnLysProAlaProAlaValProLysIleLeuAspPheSerValGly 1226

Qy     1714 ------------------------ACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAA 1749
                                    ||||||:::        :::|||     ||||||
Db     1227 ArgThrPheThrAlaIleArgGlnThrAlaIleLysLeuValPheAspThrPheLeuGlu 1246

Qy     1750 GAA------------------GCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGAC 1791
                                 |||||||||          ||||||
Db     1247 ArgAspSerProAsnAlaAlaArgGluAlaGlnGluPheGluLeuSerIleAlaLysGln 1266

Qy     1792 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 1851
                                    ||||||      :::         |||   |||
Db     1267 TyrThrAspGlyGlnLysTyrArgIleAsnIleGlyHisLysValAlaAlaLeuArgLys 1286

Qy     1852 CAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCC------AGTCTTTTA 1905
             :::      :::      :::   :::::::::       ::::::         :::|||
Db     1287 GluAsnThrSerGlyIleLeuGluValAsnLysAsnAlaValSerHisAspLysIleLeu 1306

Qy     1906 GCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCT 1965
            |||                     |||:::   |||       :::
Db     1307 AlaGlyGly---------------ProLysAspAsnCysThrValAlaArgGlyArgLys 1321

Qy     1966 AGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAA 2025
            ::::::   |||                                               |||
Db     1322 ThrHisValAsp--------------------------HisArgGlnLeuSerIleGlu 1332

Qy     2026 AAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTG 2085
            ||||||      |||          |||:::      |||    :::      |||
Db     1333 LysLeuHisProLeuLeuLeuGlnPheLysLeuThrThrSerGluLeuGluThrAsnAla 1352

Qy     2086 GTGACATTTACAAGACAGAGT--------------------------------------- 2106
                            |||
Db     1353 TyrProMetArgArgAspGlySerThrLysAlaValSerIleAlaAspThrValTyrThr 1372

Qy     2107 ------------------CTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTG 2148
                              |||:::|||:::                |||     :::
Db     1373 GlnAsnLysLysMetPheLeuAspAspTyrAspMetSerArgAsnCysSerArgCysAsn 1392

Qy     2149 ACACGCTTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAG 2208
                     ::::::         ||||||                    |||::::::
Db     1393 LysGluPheLysLeuSerProAsnGlyThr--------------------MetIleArg 1405

Qy     2209 GTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTG---------CGGGACTT 2259
                           |||:::            ||||||    :::       ||||||
Db     1406 SerThrGlyIleCysArgTyrHisAsnArgGlyValAla-IleAsnGlyLysArgAspTh 1425

Qy     2260 GTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCAC---------GG 2310
              ::::::|||:::                |||      |||      |||:::         ||
Db     1425 rPheArgLysArgTyr---------SerCysCysAsnGluGluPheAsnValAlaLeuGl 1442

Qy     2311 CTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAGTAC 2370
             ||||                                   |||        :::    |||::::::
Db     1442 yCysLysPheSerAspValHisValThrAspGlnLeuPheLysLysGluLeuSerThrPh 1462

Qy     2371 AGTGAATACACAGGCT-------------------------------------------- 2386
              |||:::|||
Db     1462 eValSerThrProValProValProAsnAspGlnArgSerThrArgValTyrAlaLeuAs 1482

Qy     2387 -------ATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATG 2425
                   |||:::      |||   ||||||          ::::::
Db     1482 pCysGluMetValTyrThrIleAlaGlyProAlaLeuAlaArgLeu 1497
```

```
RESULT 3
BIR1_SCHPO
ID   BIR1_SCHPO     STANDARD;      PRT;   997 AA.
AC   O14064; Q9USG4;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bir1 protein (Chromosome segregation protein cut17).
GN   BIR1 OR CUT17 OR PBH1 OR SPCC962.02C OR SPCP31B10.10C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21439264; PubMed=11554922;
RA   Morishita J., Matsusaka T., Goshima G., Nakamura T., Tatebe H.,
RA   Yanagida M.;
RT   "Bir1/Cut17 moving from chromosome to spindle upon the loss of
RT   cohesion is required for condensation, spindle elongation and
RT   repair.";
RL   Genes Cells 6:743-763(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99398681; PubMed=10468581;
RA   Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,
RA   Vaux D.L., Lithgow T.;
RT   "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell
RT   division.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175(1999).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=21850422; PubMed=11861551;
RA   Rajagopalan S., Balasubramanian M.K.;
RT   "Schizosaccharomyces pombe Bir1p, a nuclear protein that localizes to
RT   kinetochores and the spindle midzone, is essential for chromosome
RT   condensation and spindle elongation during mitosis.";
RL   Genetics 160:445-456(2002).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=20035862; PubMed=10571085;
RA   Rajagopalan S., Balasubramanian M.K.;
RT   "S. pombe Pbh1p: an inhibitor of apoptosis domain containing protein
RT   is essential for chromosome segregation.";
RL   FEBS Lett. 460:187-190(1999).
CC   -!- FUNCTION: Seems to act in the pleiotropic control of cell
CC       division. Has a role in chromosome segregation by recruiting
CC       condensin and ark1 kinase to appropriate sites as the cell
CC       progresses through mitosis.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Interacts with the outer
CC       centromeric regions of the chromosomes during interphase. After
```

```
CC       chromatid separation moves to the middle of the spindle.
CC   -!- SIMILARITY: Contains 2 BIR repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB031034; BAA83415.1; -.
DR   EMBL; AL031323; CAA20434.1; -.
DR   EMBL; AL121859; CAB58376.1; -.
DR   PIR; T43523; T43523.
DR   HSSP; Q13490; 1QBH.
DR   GeneDB_SPombe; SPCC962.02c; -.
DR   InterPro; IPR001370; BIR.
DR   Pfam; PF00653; BIR; 2.
DR   SMART; SM00238; BIR; 2.
DR   PROSITE; PS01282; BIR_REPEAT_1; FALSE_NEG.
DR   PROSITE; PS50143; BIR_REPEAT_2; 2.
KW   Cell division; Mitosis; Nuclear protein; Repeat.
FT   REPEAT       25     99       BIR 1.
FT   REPEAT      120    194       BIR 2.
FT   DOMAIN       80     83       POLY-ASP.
FT   DOMAIN      312    319       POLY-ASP.
FT   DOMAIN      487    490       POLY-SER.
SQ   SEQUENCE   997 AA;  112579 MW;  952A6BAFA5C489F4 CRC64;

Alignment Scores:
Pred. No.:              0.000527       Length:        997
Score:                  171.00         Matches:       188
Percent Similarity:     34.44%         Conservative:  133
Best Local Similarity:  20.17%         Mismatches:    344
Query Match:            2.52%          Indels:        267
DB:                     1              Gaps:          45

US-09-302-812-5 (1-3814) x BIR1_SCHPO (1-997)

Qy         46 TGGGAGCCC-----------------------------------------TGCACGAAA 63
              |||||||||                                         |||
Db        173 TrpGluProAspAspAspProTyrThrGluHisLysArgArgArgAlaAspCysValPhe 192

Qy         64 GCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCT 123
                    |||       :::   :::||||||               ||||||
Db        193 PheThrTrpLysAspProAsnSerLeuSerProThr---------LysLeuSerPheLeu 209

Qy        124 GGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCG 183
                                 :::||||||:::|||                :::    |||
Db        210 SerThrSerAsn------IleAspProGluAspLeuThrGluAspAsnSerIleLeuPro 227

Qy        184 TCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGACCGCACAGAGGCAACGCCACCTCGTTT 243
                 ||||||
Db        228 ValSerPro-------------------------------------------------- 230

Qy        244 GTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAA 303
                                    |||       ||||||     ||||||
Db        231 ----------------------ThrArgAspSerThrLysSerHisLysThrLeuAsn 242

Qy        304 ---TCAGAAAGTAAAGAAAACAAC---AATACAAGAATTGACTCCATGATG--------- 348
                 |||    |||:::::||||||   |||    |||       :::|||
Db        243 PheSerProSerArgLysAsnAsnLeuAsnAlaArgProLeuThrMetSerLeuTyrThr 262

Qy        349 ---AGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAATGTT 405
                 :::|||   :::||||||:::         |||   :::         :::
Db        263 AsnThrSerGluGluLysAspSer---GlnProThrArgAlaProGlnSerProThrLys 281

Qy        406 CCTCAGCTA---------------AATCTTGATAAATCACCCACAGAAAAGAGTTCACAG 450
              |||   |||                        :::|||||||||
Db        282 ProValLeuLeuThrAlaProArgArgLysAsnLysSerPro------------------ 295

Qy        451 TATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAG---------------TGGCAGAAT 495
                    ::::::          :::|||   |||                :::   :::
Db        296 ---------LysLysSerLysProAlaValPheLysProValLysProIlePheSerAsp 312

Qy        496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
              |||    :::       :::    |||    ||||||:::                |||
Db        313 GluAspGluAspAspAspLeuThrAlaSerGln---------------ProPheSer 327

Qy        556 AAGCAGCTTAGTAAT------GCTAACATTGGTCAGTCACCCCACACTGATGACCACAGT 609
              |||    :::    |||                 :::    :::          |||::::::
Db        328 LysGlyIleCysAsnAspSerMetGlnValAlaLysLysAsnPheThrGluGluIlePro 347

Qy        610 GACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCA 669
                :::    |||:::|||    :::    :::        ::::::|||
Db        348 LeuLysGluAspGluLysAspAsnGluLeuGluHisLeuValSerProAlaThrSerVal 367

Qy        670 AATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCT 729
              :::|||
Db        368 HisThr------------------------------------------------------ 369

Qy        730 CGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAG 789
                    :::|||    |||    ||||||    |||
Db        370 ------ThrValSerAspIleThrGlyHisGlnSer------------------------ 379

Qy        790 AGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAA 849
                    ::::::|||          |||
Db        380 ------ValThrAspGluSerAspGlu--------------------------------- 386

Qy        850 TTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCT 909
                       |||    :::    :::          |||       |||    ||||||:::
Db        387 ---------GlnAsnAsnCysMetSerThrProProLysIleGluIleGluSerLysIle 403

Qy        910 GAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAA 969
              |||    :::    |||    :::                   :::       |||
Db        404 GluGluGluIleSerValValSer---------------LysSerLysGluIleSerSer 418

Qy        970 GAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTT 1029
                 |||    |||       ||||||:::                |||   |||   :::
Db        419 SerValSerSerValGlyLysGluGlnAsnHisThrGluLysGlnValAlaIleGluThr 438

Qy       1030 TCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGA 1089
                 |||      |||    :::    :::
Db        439 ProGluGlnGlnLysValGluLysGluAspGlu--------------------------- 449

Qy       1090 AGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAA------AATAATGCTGGGACCAGT 1143
                             |||         :::|||               :::
Db        450 ---------------HisLeuAsnLeuGlnGlySerPheIleGluGluSerThrLysGln 464

Qy       1144 GACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCC 1203
              ::::::::::|||||||||       :::|||        :::
Db        465 ProIleSerSerLysProSerThrSerSerProAspMetThrAspAlaAlaThrGlyGly 484

Qy       1204 AAGCAGCATGGAAAA------AGGGATTCT---------------------------AAA 1230
              :::                ||||||
Db        485 ArgValSerSerSerSerPheArgAspLysIleLeuGlnThrAsnPheSerProArgSer 504

Qy       1231 ATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAA 1290
                 |||    |||       |||||||||          |||       |||:::
Db        505 ThrIleAspSerPheSerAsnIleSerLysLys------ArgAsnSerGluGluAlaAsn 522

Qy       1291 GTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCA 1350
                 :::       |||       |||||||||                       |||
Db        523 AspGluAsnAspGluThrAsnLeuLysIlePro-------------------IlePro 535

Qy       1351 GAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATC 1410
              |||||||||          :::    :::    :::    :::
Db        536 GluLysLysArgLysPheGlnGluValLeuGlnSerLysAsnIle--------------- 550

Qy       1411 CATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTG 1470
                 |||    |||          :::                |||    |||       |||
Db        551 ---LeuValSerSerThrGluAspSerHisGluProValLysValThrGluAspSerGln 569

Qy       1471 AGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAA 1530
                 |||       |||    |||                  ::::::||||||   :::||||||
```

```
Db     570 ThrAlaIleHisValSerLys---------------PheGluAspLeu---GluAsnLys 583

Qy    1531 CATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTG---GAAGATGAGAATGGT 1587
             ::::::         |||||||||:::|||       :::    |||   ||||||
Db     584 SerMetGlu---------SerGluGlnSerLeuGlnLeuLeuSerGluSerGluAsnAsp 600

Qy    1588 GAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACA 1647
           ::::::                     ||||||          |||
Db     601 AspLys------------------ProLeuIleAspLeuIleProLeuLeuAlaIleLys 614

Qy    1648 CGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCT--------- 1698
            |||       ||||||         :::|||:::      :::     |||
Db     615 ArgLysAspAsnLeuValSerGlyValLeuGluLysGlyLysSerThrSerThrSerLys 634

Qy    1699 AAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAG 1758
              |||:::|||    |||::::::||||||||     :::|||          |||
Db     635 ThrLysPheAsp---ThrSerIleValAspPheIleGluLysProLysThrGlu------ 651

Qy    1759 GCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCA 1818
                               :::||||||:::       :::     |||:::|||
Db     652 ------------IleSerGluValLeuProGluGluLysArgLysAlaIleCys------ 665

Qy    1819 AATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATG 1878
                                 ::::::       ::::::       ||||||
Db     666 ------------AspGluSerGlnThrValArgValSerIleAspArgGlyValThrLys 681

Qy    1879 TCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGG 1938
           :::::::::    ::::::|||   :::::::::
Db     682 ThrArgAsp---ValSerSerProValSerAspGlu---------------LysSerGlu 695

Qy    1939 AATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTT 1998
            |||            :::          |||    :::        :::|||    :::
Db     696 AsnValAsnHisGluGluAlaAsnSerGlyHisThrValMetAsnValHisSerSerLeu 715

Qy    1999 GAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGA 2058
            :::    :::             :::|||    :::|||::::::          |||    :::
Db     716 AspProGlnProIleValGlnProAsnGluLeuGluSer---GlySerTyrLeuLysAsp 734

Qy    2059 GTCACAGAGAAAAAA------CCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAA 2112
            :::    ::::::            :::         |||||||||            :::
Db     735 LeuProAspArgAsnValGlyAsnSerGluLysValThrPheGlnGluAspAspIleAsn 754

Qy    2113 GATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAG 2172
                                         |||      ::::::    :::    |||     |||
Db     755 Ser--------------------ProLysLeuGlnSerLysAsnAsnGlnThrValGlu 767

Qy    2173 GGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTT 2232
                  :::                    ||||||          ||||||
Db     768 ---AlaValAsnThrGluThrSerAspLysLeuGlnGluLysGluAlaAsn--------- 783

Qy    2233 GGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCT 2292
                                                           |||    |||
Db     784 ------------------------------------------HisGluLeuGluAsnIle 789

Qy    2293 GAATTGATTGTTTCACGGCTGTTCACTGAGGTG-------------------------- 2325
            |||    |||        :::|||    |||||||||
Db     790 GluLysIleGluGluLysLeu---ThrGluValAspLysValSerLeuSerAspAlaPhe 808

Qy    2326 ------------CTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGT 2373
                       :::    ::::::             :::        ||||||          |||
Db     809 ProAspGlnGluIleLysAsnSerArgThrSerValGlnAsnGlyThrArgSerValSer 828

Qy    2374 GAATACACAGGCTATGCTGAAACTTATCGTTGGG------CCCGAAGCCATGAAGATGGG 2427
            :::     |||          ||||||     |||                    |||
Db     829 LysAsnThrProGluLysGluThrLys-ValAspLysIleAspAsnValSerLysLysAs 848

Qy    2428 AGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487
             |||:::    :::    ||||||                   :::          |||:::
Db     848 pValGluThrSerProGlySerCysGluThrSerSerAlaPheAlaLysThrTyrAlaGl 868

Qy    2488 TTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547
                 :::    |||||||||:::|||   |||          ||||||:::     |||::::::::
Db     868 uLysGluValThrSerIleAsnLeu--Pro---SerValArgLysProLeuAspGluSer 886

Qy    2548 TAC---------------------------------------TGCGGATTCCTCCGTCCT 2568
            |||                                       ||||||   |||
Db     887 TyrTyrAspHisSerIleSerProPheAspProLeuCysGlnSerSerPheLeuAlaPro 906

Qy    2569 GGAGTTCCTTCTGAAAATCTTTCTGCAGTG 2598
              |||     ::::::     |||:::
Db     907 GlnThrProValLysSerLysHisAlaLeu 916
```

```
RESULT 4
NP14_HUMAN
ID   NP14_HUMAN     STANDARD;      PRT;   699 AA.
AC   Q14978; Q15030;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE   nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE   phosphoprotein 1).
GN   NOLC1 OR KIAA0035.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Leukemia;
RX   MEDLINE=95386590; PubMed=7657714;
RA   Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
RT   "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
RT   protein p130 are associated with nucleologenesis.";
RL   J. Cell Sci. 108:1911-1920(1995).
RN   [2]
RP   SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051387; PubMed=7584026;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. I.
RT   The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT   analysis of randomly sampled cDNA clones from human immature myeloid
RT   cell line KG-1.";
RL   DNA Res. 1:27-35(1994).
RN   [3]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=96205319; PubMed=8630004;
RA   Pai C.-Y., Yeh N.-H.;
RT   "Cell proliferation-dependent expression of two isoforms of the
RT   nucleolar phosphoprotein p130.";
RL   Biochem. Biophys. Res. Commun. 221:581-587(1996).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=97168979; PubMed=9016786;
RA   Chen H.-K., Yeh N.-H.;
RT   "The nucleolar phosphoprotein P130 is a GTPase/ATPase with intrinsic
RT   property to form large complexes triggered by F- and Mg2+.";
RL   Biochem. Biophys. Res. Commun. 230:370-375(1997).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=20036810; PubMed=10567578;
RA   Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;
RT   "Human Nopp140, which interacts with RNA polymerase I: implications
RT   for rRNA gene transcription and nucleolar structural organization.";
RL   Mol. Cell. Biol. 19:8536-8546(1999).
CC   -!- FUNCTION: Related to nucleologenesis, may play a role in the
CC       maintenance of the fundamental structure of the fibrillar center
CC       and dense fibrillar component in the nucleolus. It has intrinsic
CC       GTPase and ATPase activities. May play an important role in
CC       transcription catalyzed by RNA polymerase I.
CC   -!- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPA194)
CC       and with casein kinase-II.
CC   -!- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the
CC       cytoplasm. At telophase it begins to assemble into granular-like
```

```
CC       pre-nucleolar bodies which are subsequently relocated to nucleoli
CC       at the early G1-phase.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Alpha;
CC         IsoId=Q14978-1; Sequence=Displayed;
CC       Name=Beta;
CC         IsoId=Q14978-2; Sequence=VSP_004338;
CC   -!- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
CC       cycles on CK2 and PKC sites. There is evidence suggesting that
CC       CDC2 kinase phosphorylates p130 at the M-phase.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z34289; CAA84063.1; -.
DR   EMBL; D21262; BAA04803.1; -.
DR   PIR; I38073; I38073.
DR   Genew; HGNC:15608; NOLC1.
DR   GK; Q14978; -.
DR   MIM; 602394; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005730; C:nucleolus; TAS.
DR   GO; GO:0007049; P:cell cycle; TAS.
DR   GO; GO:0007067; P:mitosis; TAS.
DR   GO; GO:0006364; P:rRNA processing; TAS.
DR   InterPro; IPR006594; LisH.
DR   InterPro; IPR007718; SRP40_C.
DR   Pfam; PF05022; SRP40_C; 1.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW   Alternative splicing.
FT   DOMAIN       10     42       LISH.
FT   DOMAIN       84    566       11 X 12 AA APPROXIMATE REPEATS OF AN
FT                                ACIDIC SERINE CLUSTER.
FT   REPEAT       84     95       ACIDIC SERINE CLUSTER 1.
FT   REPEAT      125    136       ACIDIC SERINE CLUSTER 2.
FT   REPEAT      167    178       ACIDIC SERINE CLUSTER 3.
FT   REPEAT      221    232       ACIDIC SERINE CLUSTER 4.
FT   REPEAT      264    275       ACIDIC SERINE CLUSTER 5.
FT   REPEAT      325    336       ACIDIC SERINE CLUSTER 6.
FT   REPEAT      363    375       ACIDIC SERINE CLUSTER 7.
FT   REPEAT      425    436       ACIDIC SERINE CLUSTER 8.
FT   REPEAT      470    481       ACIDIC SERINE CLUSTER 9.
FT   REPEAT      519    529       ACIDIC SERINE CLUSTER 10.
FT   REPEAT      555    566       ACIDIC SERINE CLUSTER 11.
FT   DOMAIN       68     82       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      204    382       INTERACTS WITH RPA194.
FT   DOMAIN      384    587       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      601    617       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   MOD_RES     563    563       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   VARSPLIC    241    241       K -> KVWTITSVRAE (in isoform Beta).
FT                                /FTId=VSP_004338.
FT   CONFLICT      3      3       D -> A (IN REF. 2).
FT   CONFLICT    133    133       R -> S (IN REF. 2).
FT   CONFLICT    291    292       YA -> SV (IN REF. 2).
FT   CONFLICT    456    456       S -> P (IN REF. 2).
SQ   SEQUENCE   699 AA;  73720 MW;  DFD4AD94EDF659FB CRC64;

 Alignment Scores:
 Pred. No.:              0.000968      Length:        699
 Score:                  166.50        Matches:       124
 Percent Similarity:     34.58%        Conservative:  70
 Best Local Similarity:  22.10%        Mismatches:    228
 Query Match:            2.46%         Indels:        140
 DB:                     1             Gaps:          22
```

```
US-09-302-812-5 (1-3814) x NP14_HUMAN (1-699)

Qy         47 GGGAGCCCTGCACGA------------------AAGCGCGCTGGGGCGCCGCT-GGAACT 87
              |||    ||||||:::                  ||||||    |||     |||     |||
Db         99 GlyProProAlaLysLysAlaAlaValProAlaLysArgValGlyLeuProProGlyLys 118

Qy         88 TCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGT--------- 138
              :::|||          ||||||:::|||      |||                   :::     |||
Db        119 AlaAlaAlaLysAlaSerGluSerSerSer------SerGluGluSerArgAspAspAsp 136

Qy        139 ---GTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCC 195
                 |||   |||       |||||||||     |||   |||   :::   |||
Db        137 AspGluGluAspGlnLysLysGlnProValGlnLysGlyValLysProGlnAlaLysAla 156

Qy        196 TGCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTC------ 249
                                 :::|||   |||   :::      ::::::
Db        157 ---------AlaLysAlaProProLysLysAlaLysSerSerAspSerAspSerAspSer 173

Qy        250 -----------------------------------AAACAAAAGACTATTACTACTTGG 273
                                                 |||         :::|||
Db        174 SerSerGluAspGluProProLysAsnGlnLysProLysIleThrProValThrValLys 193

Qy        274 ATGGATACTAAAGGA---CCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACA 330
                    ||||||      ||||||    |||    :::           |||        |||
Db        194 AlaGlnThrLysAlaProProLysProAlaArgAlaAlaProLysIleAlaAsnGlyLys 213

Qy        331 AGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAA 390
                    |||      ||||||             :::               :::   |||
Db        214 AlaAlaSerSerSerSerSerSerSerSerSerSerSerSerAspAspSerGluGluGlu 233

Qy        391 AAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAA------------------------ 426
              |||              |||:::        :::   |||
Db        234 LysAlaAlaAlaThrProLysLysThrValProLysLysGlnValValAlaLysAlaPro 253

Qy        427 ---------------TCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACT 471
                             :::||||||   |||||||||            :::   :::   :::
Db        254 ValLysAlaAlaThrThrProThrArgLysSerSer---------SerSerGluAspSer 270

Qy        472 GCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAG 531
              :::|||                  :::|||   :::      ::::::   :::   ::::::
Db        271 SerSer---------------AspGluGluGluGluGlnLysLysProMetLysAsnLys 285

Qy        532 ---------------------CCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAAT 570
                                   ||||||          |||      ||||||:::
Db        286 ProGlyProTyrSerTyrAlaProProProSerAlaProProProLysLys--------- 302

Qy        571 GCTAACATTGGTCAGTCACCC--------------------------------------- 591
                 ::::::|||      |||
Db        303 ---SerLeuGlyThrGlnProProLysLysAlaValGluLysGlnGlnProValGluSer 321

Qy        592 ------CACACTGATGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAG 645
                    :::      |||      ||||||:::            ||||||::::::
Db        322 SerGluAspSerSerAspGluSerAspSerSerSerGluGluGluLysLys--------- 338

Qy        646 TTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGA 705
                        |||   |||         :::|||   |||                  |||:::
Db        339 ------ProProThrLysAlaValValSerLysAlaThrThrLysProProProAlaLys 356

Qy        706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765
                                   :::   :::|||      |||      :::
Db        357 Lys------------------AlaAlaGluSerSerAspSerSerAspSerAspSer 370

Qy        766 GAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGA 825
                 |||   |||         |||         |||                        |||
Db        371 SerGluAspAspGluAlaProSerLysPro------------------------AlaGly 382

Qy        826 ACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCT 885
              |||         :::   ||||||                  :::               ||||||
Db        383 ThrThrLysAsnSerSerAsnLys------------ProAlaValThrThrLysSerPro 398

Qy        886 CCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATG------------------------ 921
                    :::      :::   |||:::   |||:::
```

```
Db        399 AlaValLysProAlaAlaAlaProLysGlnProValGlyGlyGlyGlnLysLeuLeuThr 418

Qy        922 ---------GATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAA 972
                       |||      :::|||                  ||||||   ::::::
Db        419 ArgLysAlaAspSerSerSerSerGluGluGluSerSerSerSerGluGluGluLysThr 438

Qy        973 ACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCA 1032
                       |||        :::        :::::::::  |||      |||
Db        439 LysLysMetValAlaThrThrLysProLysAlaThrAlaLysAlaAlaLeuSerLeuPro 458

Qy       1033 AGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGT 1092
               :::   ||||||   :::        |||               :::     |||
Db        459 AlaLysGlnAlaProGlnGlySerArgAspSerSerSerAspSerAspSerSerSerSer 478

Qy       1093 GAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAAT 1152
              |||                  |||   |||            |||   :::
Db        479 Glu--------------------GluGluGluGluLysThrSerLysSerAlaValLys 491

Qy       1153 GCCAAGCCA------------TCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
                 ||||||              |||   :::   |||                 :::
Db        492 LysLysProGlnLysValAlaGlyGlyAlaAlaProSerLysProAlaSerAlaLysLys 511

Qy       1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
                 |||                 |||         |||                |||
Db        512 GlyLysAlaGluSerSerAsnSerSerSerSerAspAsp---------------SerSer 526

Qy       1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
                 |||:::  |||||||||          |||          |||       :::
Db        527 GluGluGlu-GluGluLysLeuLysGlyLysGlySerProArgProGlnAlaProLysAl 546

Qy       1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
                    ||||||   ||||||:::              :::   |||   ||||||
Db        546 aAsnGlyThrSerAlaLeuThrAla-GlnAsnGlyLysAlaAlaLysAsnSerGluGluG 566

Qy       1381 GAAATGAG------AAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCT 1434
                 |||      |||   |||:::|||                            :::
Db        566 luGluGluGluLysLysLysAlaAlaVal-----------------------ValValS 578

Qy       1435 GCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACC 1493
                 ||||||   |||                 :::   :::|||:::   :::   :::||||||
Db        578 erLysSerGlySerLeuLysLysArgLysGlnAsnGluAlaAlaLysGluAlaGluThr 597
```

```
RESULT 5
AF4_HUMAN
ID   AF4_HUMAN      STANDARD;      PRT;  1210 AA.
AC   P51825;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   AF-4 protein (Proto-oncogene AF4) (FEL protein).
GN   MLLT2 OR AF4 OR FEL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281633; PubMed=8506309;
RA   Nakamura T., Alder H., Gu Y., Prasad R., Canaani O., Kamada N.,
RA   Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
RA   Canaani E.;
RT   "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT   acute leukemia share sequence homology and/or common motifs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93184301; PubMed=8443374;
RA   Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
RA   Cleary M.L.;
RT   "A serine/proline-rich protein is fused to HRX in t(4;11) acute
RT   leukemias.";
RL   Blood 81:1124-1131(1993).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DISEASE: Involved in acute leukemias through a chromosomal
CC       translocation t(4;11)(q21;q23) that involves MLLT2 and MLL/HRX.
CC       The result is a rogue activator protein.
CC   -!- SIMILARITY: Belongs to the AF4 family.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/AF4.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L13773; AAA58360.1; -.
DR   EMBL; L25050; AAA36642.1; -.
DR   PIR; A58198; A58198.
DR   PIR; I39410; I39410.
DR   Genew; HGNC:7135; MLLT2.
DR   MIM; 159557; -.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   InterPro; IPR007797; AF-4.
DR   Pfam; PF05110; AF-4; 1.
KW   Nuclear protein; Chromosomal translocation; Proto-oncogene.
FT   DOMAIN      483    492       POLY-SER.
FT   DOMAIN      835    843       POLY-SER.
FT   DOMAIN      866    869       POLY-PRO.
FT   DOMAIN      871    874       POLY-SER.
FT   CONFLICT     46     46       K -> R (IN REF. 2).
FT   CONFLICT    624    624       E -> G (IN REF. 2).
FT   CONFLICT    899    905       SASSTKS -> VPAVPRV (IN REF. 2).
FT   CONFLICT    928    929       EH -> AD (IN REF. 2).
FT   CONFLICT    999    999       I -> N (IN REF. 2).
FT   CONFLICT   1096   1096       A -> AR (IN REF. 2).
FT   CONFLICT   1140   1140       N -> I (IN REF. 2).
FT   CONFLICT   1177   1210       STNVCTLALNSSLVDLVHYTRQGFQQLQELTKTP -> RQ
FT                                MCAPWPSTAVWWTWCTIHDRVFSSYKN (IN REF. 2).
SQ   SEQUENCE   1210 AA;  131421 MW;  F0E334DF8FC2FF04 CRC64;

Alignment Scores:
Pred. No.:              0.00125        Length:          1210
Score:                  166.00         Matches:         216
Percent Similarity:     33.43%         Conservative:    140
Best Local Similarity:  20.28%         Mismatches:      402
Query Match:            2.45%          Indels:          307
DB:                     1              Gaps:            48

US-09-302-812-5 (1-3814) x AF4_HUMAN (1-1210)

Qy         15 TGCGGG---------TCCCAGCATGAGTGCGGGCCCCGGCTG---------------GGA 50
              ||||||         |||||||||            ||||||               |||
Db        151 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 170

Qy         51 GCCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTC 110
              :::   |||               ||||||                  ||||||   ::
Db        171 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 190

Qy        111 CCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTT 170
              :  |||   |||   |||:::         |||         :::|||
Db        190 rAspSerAlaProGluArgGluLeuSerProLeuIleSerLeuProSerPro-------- 207

Qy        171 CAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAA 230
                  |||||||||   ||||||         :::   :::      |||
Db        208 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnGl 226

Qy        231 CGCCACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACC 290
              ::::::                                   :::     |||   ||
Db        226 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 246

Qy        291 CAAG------ACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCAT 344
              ||||            :::         ||||||         |||   ||||||::
Db        246 oLysAspLeuAlaValLysValHisAspLysGlu---------ThrProGlnAspSerLe 263
```

```
Qy   345 GATGAGTTCTGTGCAGAAAGATAAC------------------TTTTACCCACATAAGGT 386
Db   263 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLysSe 283
Qy   387 GGAAAAATTGGAAAATGTTCCTCAGCTAAATCTT------------------GATAAATC 428
Db   283 rValAlaMetGlnGlnLysProThrAlaTyrValArgProMetAspGlyGlnAspGlnAl 303
Qy   429 ACCCACAGAAAAG------------------AGTTCACAGTATTTGAACCAACAGCAGAC 470
Db   303 aProSerGluSerProGluLeuLysProLeuProGluAspTyr------ArgGlnGlnTh 321
Qy   471 T---------------------------------GCGAGTGTGTGCAAGTGGCAGAATGA 497
Db   321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341
Qy   498 AGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAA 557
Db   341 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluIleLeuLy 361
Qy   558 GCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGA 617
Db   361 sGluMetThrHisSer------TrpProProProLeuThrAlaIleHisThrProSerTh 379
Qy   618 TCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAA 677
Db   379 rAlaGluProSerLys------------PheProPheProThrLysAspSerGlnHisVa 395
Qy   678 GCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTC 737
Db   395 lSerSerValThrGlnAsnGlnLysGlnTyrAsp--------------------ThrSe 408
Qy   738 TGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGT---CC 794
Db   408 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 428
Qy   795 TTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGAC 854
Db   428 nLeuSerAsp------SerGluAsp------------------SerAspSerGluGlnTh 440
Qy   855 TGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTC 914
Db   440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu-- 459
Qy   915 ACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAAC 974
Db   460 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 479
Qy   975 AAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAG 1034
Db   479 rSerSerAspSerGluSerGluSerSerSerSerAspSerGluGluAsnGluProLeuGl 499
Qy  1035 TTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGG---------------TATTT 1079
Db   499 uThrProAlaProGluProGluProProThrThrAsnLysTrpGlnLeuAspAsnTrpLe 519
Qy  1080 GACTAAGGGAAGTGAA---------------------------------------GTTAG 1100
Db   519 uThrLysValSerGlnProAlaAlaProProGluGlyProArgSerThrGluProProAr 539
Qy  1101 ATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAG-- 1158
Db   539 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 559
Qy  1159 -------CCATCTGGAAACTCTTCTAGCCTT---------------------------- 1182
Db   559 rLysAspProProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 579
Qy  1183 -------AATGTAGAGTGTAGAAGTTCC------------------------------- 1203
Db   579 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 599
Qy  1204 ----------AAGCAG-------------------------------------------- 1209
Db   599 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 619
Qy  1210 -------CATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTC 1262
Db   619 rSerLeuGlnGlyGluArgGluProGlyLeuLeuProTyrGlySerArgAspGlnThrSe 639
Qy  1263 AGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCC 1322
Db   639 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaAlaSerAsnGluPr 659
Qy  1323 AAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGA 1382
Db   659 oLys---------ProAlaValProProSerSerGlu--------------------- 668
Qy  1383 AATGAGAAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCA 1442
Db   669 ----LysLysLysHisLysSerSerLeuProAlaProSer-------------------- 680
Qy  1443 CACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAC 1502
Db   681 ---------------------------LysAlaLeuSerGlyProGluProAlaLysAs 691
Qy  1503 ACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTT 1562
Db   691 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 710
Qy  1563 GTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCA 1622
Db   711 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValValVa 730
Qy  1623 GACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATA 1682
Db   730 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----------- 746
Qy  1683 CAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGT 1742
Db   747 --------------------------------------------------------LysLe 748
Qy  1743 ACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAAT 1802
Db   748 uLeuSerProLeuArgAspThrProProProGlnSerLeu---------MetValLysIl 765
Qy  1803 TGCACTCTGTCTGCCAAATATTTGCACCCAGCCA------------------------- 1836
Db   765 eThrLeuAspLeuLeuSerArgIleProGlnProProGlyLysGlySerArgGlnArgLy 785
Qy  1836 ------------------------------------------------------------ 1836
Db   785 sAlaGluAspLysGlnProProAlaGlyLysLysHisSerSerGluLysArgSerSerAs 805
Qy  1837 -------ATACCACTCCTGAAACAGAAGATG----------------------AATCA 1865
Db   805 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 825
Qy  1866 TTCTGTCACGATG---------------TCACAGGAACAGATCGCCAGTCTTTTAGCTAA 1910
Db   825 sLysIleArgLeuGluLysGluIleLysSerGlnSerSerSerSerSerSerHisLy 845
Qy  1911 TGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTA 1970
Db   845 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 865
Qy  1971 CCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAA-- 2028
Db   865 uProProProProValSer---------------SerSerSerGlnLysProAlaLysPr 880
Qy  2029 ----CTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAG---------AAAAAACC 2075
Db   880 oAlaLeuLysArg------------SerArgArgGluAlaAspThrCysGlyGlnAspPr 896
Qy  2076 TACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGAT------TTTCCAGAATGGGA 2129
```

```
                |          :::   |||:::   :::   :::|||        |||:::
Db         896 oProLysSerAlaSerSerThrLysSerAsnHisLysAspSerSerIleProLysGlnAr 916

Qy        2130 AAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATA-------- 2181
                |||  |||        :::|||          ::::::|||:::
Db         916 gArgValGluGlyLysGlySerArgSerSerSerGluHisLysGlySerSerGlyAspTh 936

Qy        2182 ------------------------------GAAGGCAACGGCCGA---GGCATGCTACA 2207
                                              ||||||   :::   |||      ||
Db         936 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 956

Qy        2208 GGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTT---GTACA 2264
                ||||  |||       :::                          |||  |||  :::::
Db         956 nValLysPheAspLysGln--------------------GlnAlaAspLeuHisMetAr 969

Qy        2265 AGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTG----------- 2313
                :|||       :::              ||||||:::      |||:::
Db         969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaPh 989

Qy        2314 -------TTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACA 2366
                               ||||||         ||||||   |||   |||   :::|||
Db         989 eLysTyrLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGluSe 1009

Qy        2367 GTACAGTGAA---------TACACAGGCTATGCTGAAACT------------------TA 2399
                   |||            |||:::  |||:::||||||
Db        1009 rGlnSerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIl 1029

Qy        2400 TCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAG----------- 2448
                       :::|||      |||   :::                :::
Db        1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaVa 1049

Qy        2449 -------CGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCT 2501
                       ||||||     |||          ::::::         ||||||
Db        1049 lLeuCysMetArgCysGlnSerIle------LeuAsnMetAlaMetPheArgCysLysLy 1067

Qy        2502 CGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCT 2561
                |||       :::     |||      |||    |||||||||   :::
Db        1067 sAsp------IleAlaIleLysTyrSerArgThrLeuAsnLysHisPhe----------- 1081

Qy        2562 CCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGG 2621
                                  |||:::   |||   ||||||            |||
Db        1082 ------------------GluSerSerSerLysValAlaGlnAlaProSerProCysIl 1095

Qy        2622 TGCCTTTGGGGGT 2634
                  |||     |||
Db        1095 eAlaSerThrGly 1099
```

RESULT 6

AKA6_RAT

```
ID   AKA6_RAT       STANDARD;      PRT;  2314 AA.
AC   Q9WVC7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   A-kinase anchor protein 6 (Protein kinase A anchoring protein 6)
DE   (PRKA6) (mAKAP).
GN   AKAP6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=99343692; PubMed=10413680;
RA   Kapiloff M.S., Shillace R.V., Westphal A.M., Scott J.D.;
RT   "mAKAP: an A-kinase anchoring protein targeted to the nuclear membrane
RT   of differentiated myocytes.";
RL   J. Cell Sci. 112:2725-2736(1999).
CC   -!- FUNCTION: Binds to type II regulatory subunits of protein kinase A
CC       and anchors/targets them to the nuclear membrane or sarcoplasmic
CC       reticulum. May act as an adapter for assembling multiprotein
CC       complexes (By similarity).
CC   -!- SUBUNIT: Interacts with RII subunit of PKA, phosphatase 2B
CC       (calcineurin) and AKAP79 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Sarcoplasmic reticulum and nuclear membrane
CC       in heart muscle. Participation of multiple targeting signals allow
CC       correct intracellular targeting. These may be repeated motifs rich
CC       in basic and hydrophobic amino acids, palmitoylated/myristoylated
CC       motifs or alternatively splice targeting sequences.
CC   -!- DOMAIN: RII-alpha binding site, predicted to form an amphipathic
CC       helix, could participate in protein-protein interactions with a
CC       complementary surface on the R-subunit dimer.
CC   -!- SIMILARITY: Contains 2 spectrin repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF139518; AAD39150.1; -.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00435; spectrin; 3.
DR   SMART; SM00150; SPEC; 3.
KW   Repeat.
FT   REPEAT      768    847       SPECTRIN 1.
FT   REPEAT     1033   1148       SPECTRIN 2.
FT   DOMAIN      516    606       SER-RICH.
FT   DOMAIN     1558   1699       SER-RICH.
FT   DOMAIN     2062   2075       PKA-RII SUBUNIT BINDING DOMAIN.
SQ   SEQUENCE   2314 AA;  254349 MW;  2FB01DE0E2DDFA37 CRC64;

Alignment Scores:
Pred. No.:              0.0018        Length:          2314
Score:                  165.00        Matches:         145
Percent Similarity:     33.15%        Conservative:    90
Best Local Similarity:  20.45%        Mismatches:      265
Query Match:            2.44%         Indels:          209
DB:                     1             Gaps:            32

US-09-302-812-5 (1-3814) x AKA6_RAT (1-2314)

Qy         412 CTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACT 471
               ||||||   |||      |||   :::                  |||   ::::::
Db        1154 LeuAsnAlaAspHisGlnProMetGlnLeuIleIleValAsnLeuGluArgArgTrpGlu 1173

Qy         472 GCGAGTGTG---------TGCAAGTGGCAGAATGAA---------------GGGAAGCAT 507
               |||   |||            :::||||||                  ||||||
Db        1174 AlaIleValMetGlnAlaValGlnTrpGlnThrArgLeuGlnLysLysMetGlyLysGlu 1193

Qy         508 GCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGT 567
               :::|||   |||         :::|||                          |||
Db        1194 SerGluThrLeuAsnValIleAspPro-----------------------GlyLeuMet 1205

Qy         568 AATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGAT--------- 618
               :::    |||   |||   |||             :::      :::||||||
Db        1206 AspLeuAsn---GlyMetSerGluAspAlaLeuGluTrpAspGluThrAspIleSerAsn 1224

Qy         619 ---------------CATGAAGAA------------GACAGAGACAATCAGCAGTTTCTT 651
                              |||||||||             |||:::|||   :::      |||
Db        1225 LysLeuIleSerValHisGluGluSerAsnAspLeuAspGlnAspProGluProMetLeu 1244

Qy         652 ACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAAC 711
                  :::||||||     |||                                    :::
Db        1245 ProAlaValLysLeuGluGluThr---------------------------------His 1253

Qy         712 TGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAG 771
                 |||   ||||||        :::        ||||||   |||
Db        1254 HisLysAspSerGlyTyrGluGluGluAlaGlyAspCysGlyGlySerProTyrThrSer 1273

Qy         772 GTGGATGTGCTACCAGAGAGTCCT--------------------TTGTCAGATGTTGGT 810
                     |||   ||||||                        |||      :::|||
Db        1274 AsnIleThrAlaProSerSerProHisIleTyrGlnValTyrSerLeuHisAsnValGlu 1293
```

```
Qy    811 GCCGAGGACATTGGAACTGGACCA------AAAAATGACAACAAATTGACTGGACAAGAA 864
                 :::             |||       |||:::       |||   ||||||
Db   1294 LeuHisGluAspSerHisThrProPheLeuLysSerSerProLysPheThrGlyThrThr 1313

Qy    865 AGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGT--------------------- 903
                      :::                  |||:::|||
Db   1314 GlnProThrValLeuThrLysSerLeuSerLysAspSerSerPheSerSerThrLysSer 1333

Qy    904 --------------------------------GAGCCTGAGTCACCAATGGATGTAGAC 930
                                            |||    |||          |||
Db   1334 LeuProAspLeuLeuGlyGlySerGlyLeuValArgProTyrSerCysHisSerGlyAsp 1353

Qy    931 AACTCGAGAAACAGTTGTCAAGAT------------TCAGAAGCAGATGAAGAA------ 972
             |||:::||||||       :::           ||||||   |||   |||
Db   1354 LeuSerGlnAsnSerGlySerGluSerGlyIleValSerGluGlyAspAsnGluMetPro 1373

Qy    973 ---------ACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAAT 1023
                           :::|||          |||         |||
Db   1374 ThrAsnSerAspMetSerLeuPheSerMetValAspGlySerProSerAsnProGluThr 1393

Qy   1024 AAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083
          :::             |||      :::|||        |||     :::::::::
Db   1394 GluHisProAspProGlnMetGlyAspAlaAlaAsnValLeuGluGlnLysPheLysAsp 1413

Qy   1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGT 1143
             |||      ::::::|||                               :::   |||
Db   1414 AsnGlyGluSerIleLysLeu---------------------------SerSerValSer 1424

Qy   1144 GACTTAAATGCCAAGCCA------------TCTGGAAACTCTTCTAGCCTTAAT------ 1185
                :::      |||             :::|||   :::      ||||||
Db   1425 ArgAlaSerValSerProValGlyCysValAsnGlyLysAlaGlyAspLeuAsnSerVal 1444

Qy   1186 ------------GTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGAT--------- 1224
                            :::|||      :::         ||||||   |||
Db   1445 ThrLysHisThrAlaAspCysLeuGlyGluGluLeuGlnGlyLysHisAspValPheThr 1464

Qy   1225 --------------------------TCTAAAATT---ACAGATCATTTCATGAGAATT 1254
                                    ||||||:::            |||:::
Db   1465 PheTyrAspTyrSerTyrLeuGlnGlySerLysLeuLysLeuProMetIleMetLysGln 1484

Qy   1255 TCCAAGTCAGAGGACAGAAGA-----------------------------------AAA 1278
             :::||||||                                          |||
Db   1485 ProGlnSerGluLysAlaHisValGluAspProLeuLeuGlyGlyPheTyrPheAspLys 1504

Qy   1279 GAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCT 1338
          :::     |||:::   :::||||||   :::|||    :::                |||
Db   1505 LysSerCysLysAlaLysHisGlnAlaSerGluSerGln------------------Pro 1518

Qy   1339 AACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCT 1398
          :::     ||||||    ::::::    |||      :::      ||||||    :::
Db   1519 AspAlaProProHisGluArgIleLeuAlaSerAlaProHisGluMetGlyArgSerAla 1538

Qy   1399 CGG--------------------------TGTGGGATCCATTTGCCTTCCTTAAGACCA 1431
                                              ||||||
Db   1539 TyrLysSerSerAspIleGluLysThrPheThrGlyIleGlnSerAlaArgGlnLeuSer 1558

Qy   1432 TCTGCAAGTCACACAGTGACTGTTCGG-------------------------------- 1458
                   |||    :::    :::|||
Db   1559 LeuLeuSerArgSerSerSerValGluSerLeuSerProGlyGlyAspLeuPheGlyLeu 1578

Qy   1459 ------------------GTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
                            |||   |||:::               :::    :::
Db   1579 GlyIlePheLysAsnGlySerAspSerLeuGlnArgSerThrSerLeuGluSerTrpLeu 1598

Qy   1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAAC 1560
          |||   ||||||                                  |||   ::::::
Db   1599 ThrSerTyrLys------------------------------------SerAsnGluAsp 1606

Qy   1561 TTGTACCCTGTGGAAGATGAGAATGGT---GAGCGAACTGCAGGGAGTAGGTGGGAGCTC 1617
          |||:::                             ::::::||||||      ||||||
Db   1607 LeuPheSerCysHisSerSerGlyAspIleSerValSerSerGlySerValGlyGluLeu 1626

Qy   1618 ATTCAGACTGCA------CTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCG 1671
            :::                |||||||||:::           |||:::      :::
Db   1627 SerLysArgThrLeuAspLeuLeuAsnArgLeuGluAsnIleGlnSerProSerGluGln 1646

Qy   1672 ATTCTGAAATACAATGTG------------------GCATATTCTAAGAAATGGGACTTT 1713
            :::|||    :::|||                   :::    |||:::|||    |||
Db   1647 LysIleLysArgSerValSerAspMetThrLeuGlnSerSerSerGlnLysMetProPhe 1666

Qy   1714 ACAGCT------TTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT 1767
                     :::|||         :::     |||::::::    |||
Db   1667 AlaGlyGlnMetSerLeuAspValAlaSerSerIleAsnGluAspSerProAlaSerLeu 1686

Qy   1768 TTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGC 1827
                   |||         ::::::                            ::::::|||
Db   1687 ThrGluLeuSerSerSerAspGluLeu------------------------SerLeuCys 1698

Qy   1828 ACCCAGCCAATACCACTCCTGAAACAGAAG------------------------------ 1857
          ::::::   |||   |||   |||   |||
Db   1699 SerGluAspIleValLeuHisLysAsnLysIleProGluSerAsnAlaSerPheArgLys 1718

Qy   1858 ---ATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCT 1914
             :::|||   ||||||          :::            |||::::::   |||
Db   1719 ArgLeuAsnArgSerValAlaAspGluSerAspValAsnValSerMetIleValAsnVal 1738

Qy   1915 TTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCA 1974
                ||||||          :::   :::         |||:::  :::|||    :::
Db   1739 ---SerCysThrSerAlaCysThrAspAspGluAspAspSerAsp-LeuLeuSerSerSe 1757

Qy   1975 GACATTAACTTCAATCGGT---------TGTTTGAAGGACGTTCATCAAGGAAACCAGAA 2025
            |||||||||               ||||||||||||     :::    :::
Db   1757 rThrLeuThrLeuThrGluGluGluLeuCysLeuLysAsp------GluAspAspAspSe 1775

Qy   2026 AAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAG-------------------- 2065
            :::                   :::|||||||||
Db   1775 rSerIleAlaThrAspAspGluIleTyrGluGluSerAsnLeuMetSerGlyLeuAspTy 1795

Qy   2066 ----AGAAAAAACCTACAGGATTGG 2086
              :::      ||||||   |||
Db   1795 rIleLysAsnGluLeuGlnThrTrp 1803
```

RESULT 7
PCLO_MOUSE
ID   PCLO_MOUSE   STANDARD;   PRT;   5038 AA.
AC   Q9QYX7; Q9QYX6; Q9QZJ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-
DE   derived HLMN protein).
GN   PCLO OR ACZ.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RP   TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   REVISIONS.
RA   Kilimann M.W.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 4502-4682 FROM N.A.
RC   TISSUE=Brain;
RA   Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;

```
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   INTERACTION WITH RIMS2.
RX   MEDLINE=22384373; PubMed=12401793;
RA   Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA   Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT   "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT   cAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis.";
RL   J. Biol. Chem. 277:50497-50502(2002).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking.
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin.
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9QYX7-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC       stomach. Not detected in other tissues analyzed including adrenal
CC       gland, testis and pancreas.
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19185; CAB60731.2; -.
DR   EMBL; Y19186; CAB60732.2; -.
DR   EMBL; AF181269; AAD55786.2; -.
DR   HSSP; P04410; 1A25.
DR   MGD; MGI:1349390; Pclo.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; IDA.
DR   GO; GO:0019933; P:cAMP-mediated signaling; IDA.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
DR   GO; GO:0030073; P:insulin secretion; IDA.
DR   GO; GO:0017157; P:regulation of exocytosis; IDA.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      371    470       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     502    526       C4-TYPE (POTENTIAL).
FT   ZN_FING     967    990       C4-TYPE (POTENTIAL).
FT   DOMAIN     2305   2329       POLY-PRO.
FT   DOMAIN     4394   4488       PDZ.
FT   DOMAIN     4607   4705       C2 DOMAIN 1.
FT   DOMAIN     4922   5012       C2 DOMAIN 2.
FT   VARSPLIC   4829   4833       TKPTN -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003928.
FT   VARSPLIC   4834   5038       Missing (in isoform 2).
FT                                /FTId=VSP_003929.
SQ   SEQUENCE   5038 AA;  547600 MW;  DADA460CF3B40888 CRC64;

Alignment Scores:
Pred. No.:              0.00372      Length:          5038
Score:                  162.00       Matches:         122
Percent Similarity:     33.93%       Conservative:    86
Best Local Similarity:  19.90%       Mismatches:      207
Query Match:            2.39%        Indels:          198
DB:                     1            Gaps:            28

US-09-302-812-5 (1-3814) x PCLO_MOUSE (1-5038)

Qy        250 AAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAA 309
              :::|||||||||   |||          :::|||:::              :::|||   |||
Db       1355 GlnGlnLysThrGlnThrPro---SerGluThrArgAspIleSerIleSerGluGluGlu 1373

Qy        310 AGTAAAGAAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAAC 369
                 ||||||:::         ::::::   |||          |||   |||
Db       1374 IleLysGluSerGlnGluLysLysValThrSerLysLysAspSerAlaGlnGly------ 1391

Qy        370 TTTTACCCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCA 429
                 :::|||    :::      |||      ||||||   |||:::|||         |||   |||
Db       1392 ---PheProSerArgLysGluHisLysGluAsn---ProGluLeuValAspAspLeuSer 1409

Qy        430 CCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGG 489
              |||         |||         :::
Db       1410 ProArgArgAlaSerTyrAspSerValGluAsp--------------------------- 1420

Qy        490 CAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCG 549
                                                   |||            :::|||
Db       1421 -----------------------------------SerSerGluSerGluAsnSerPro 1428

Qy        550 CTACCAAAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGT 609
              :::    ::::::            :::||||||   |||         :::::::::::
Db       1429 ValAlaArgArgLysArgArgThrSerIleGlySerSer---SerSerGluGluTyrLys 1447

Qy        610 GACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCA 669
                    |||    :::      :::::::::   |||:::
Db       1448 GlnGluAspSerGlnGlySerGlyGluAspGluAspPheIle------------------ 1461

Qy        670 AATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCT 729

Db       1461 ------------------------------------------------------------ 1461

Qy        730 CGCCAGTCTGTG------------------AAAGACTGTACAGGCTGTCAACAGGAGGAG 771
              |||:::    :::                    :::|||      :::|||    :::    ||||||
Db       1462 ArgLysGlnIleIleGluMetSerAlaAspGluAspAlaSerGlySerGluAspGluGlu 1481

Qy        772 GTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGA 831
                          |||    |||    ::::::|||         :::
Db       1482 ---------PheIleArgSerGlnLeuLysGluIleGlyGly------ValThrGluSer 1496

Qy        832 CCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTT 891
                 |||    :::                    :::|||      |||
Db       1497 GlnLysArgGluGluThrLysGlyLysGlyLysSerProAlaGlyLysHisArgArgLeu 1516

Qy        892 GAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAA 951
                 ::::::|||          :::      |||   |||      |||
Db       1517 ThrArgLysSer------SerThrSerPheAspAspAspAlaGlyArgArgHisSerTrp 1534

Qy        952 GATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCC 1008
                 |||   ||||||               |||||||||         :::   :::
Db       1535 HisAspGluAspAspGlu------------ThrPheAspGluSerProGluLeuLysPhe 1550

Qy       1009 TCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGG 1068
                 :::|||    ::::::   |||                           |||   ||||||
Db       1551 ArgGluThrLysSerGlnGluSerGluGluLeuValValAlaGlyGlyGlyGlyLeuArg 1570

Qy       1069 AAA-----------------------------------CGGTATTTGACTAAGGGAAGT 1092
```

```
                 :::                                      :::|||      :::   |||
Db      1571 ArgPheLysThrIleGluLeuAsnSerThrValThrAspLysTyrSerAlaGluSerSer 1590

Qy      1093 GAAGTTAGA------TTGCATTTCCAATTTGAAGGAGAA-------------------- 1125
              :::   :::      |||:::|||        |||   |||
Db      1591 GlnLysLysThrThrLeuTyrPheAspGluGluProGluLeuGluMetGluSerLeuThr 1610

Qy      1126 ---------------------AATAATGCTGGGACCAGTGACTTAAATGCCAAG------ 1158
                                  |||:::|||   |||:::|||
Db      1611 AspSerProGluAspArgSerArgGlyGluGlySerSerSerLeuHisAlaSerSerPhe 1630

Qy      1159 CCATCTGGAAACTCT------------TCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAG 1206
                   |||   |||            |||||||||:::    :::      ||||||
Db      1631 ThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSerAspSerSerPro 1650

Qy      1207 CAGCATGGAAAAAGGGATTCTAAA------------------------------------ 1230
                |||   |||   :::||||||
Db      1651 SerHisLysLysGlyGluSerLysGlnGlnArgLysAlaArgHisArgSerHisGlyPro 1670

Qy      1231 ------------ATTACAGAT--------------------------------------- 1239
                         |||   |||
Db      1671 LeuLeuProThrIleGluAspSerSerGluGluGluGluLeuArgGluGluGluGluLeu 1690

Qy      1240 ---------------------------------CATTTCATGAGAATTTCCAAGTCAGAG 1266
                                              |||   |||   |||:::
Db      1691 LeuLysGluGlnGluLysGlnArgGluLeuGluGlnGlnGlnArgLysSerSerSerLys 1710

Qy      1267 GACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAA 1326
                :::|||::::::   |||:::|||   ||||||      |||:::   ||||||
Db      1711 LysSerLysLysAspLysAspGluLeuArgAlaGlnArg---ArgArgGluArgProLys 1729

Qy      1327 TACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATG 1386
                |||   ||||||   |||   :::            :::|||   ||||||:::
Db      1730 ThrProProSerAsnLeuSerProIleGluAsp------AlaSerProThrGluGluLeu 1747

Qy      1387 AGAAAAATGCCTCGG-----------------------TGTGGGATCCATTTGCCTTCC 1422
             |||:::                                |||      :::   ||||||
Db      1748 ArgGlnAlaAlaGluMetGluGluLeuHisArgSerSerCysSerGluTyrSerProSer 1767

Qy      1423 TTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAG 1482
             :::
Db      1768 IleGluSerAsp------------------------------------------------ 1771

Qy      1483 GTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGAT---AACAAACATGTGAAA 1539
                |||:::   |||         :::   :::   :::      |||      |||
Db      1772 ---ProGluGlyPheGluIleSerProGluLysIleIleGluValGlnLysValTyrLys 1790

Qy      1540 ATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCA 1599
             :::|||   :::      :::||||||         ||||||
Db      1791 LeuPro---ThrAlaValSerLeuTyrSerProThrAspGluGln--------------- 1804

Qy      1600 GGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAAC 1659
                                      :::::::::   |||             :::
Db      1805 ---------------------------SerValMetGlnLysGluGlyAlaGlnLysAla 1815

Qy      1660 TTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1719
             ||||||   |||         |||            |||         :::   |||
Db      1816 LeuLysSerAlaGluGluMetTyrGluGluMetMetHisLysProHisLysTyrLysAla 1835

Qy      1720 TTG------GTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAG--------------- 1758
                        :::      |||:::|||   |||:::
Db      1836 PheProAlaAlaAsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyMet 1855

Qy      1759 ------GCCCAACATTTATATCAGTCCATTTTACCTGAC 1791
                   ::::::|||:::|||::::::   |||
Db      1856 LeuIleGluAspTyrIleTyrGluSerLeuValGluAsp 1868
```

```
RESULT 8
PCLO_HUMAN
ID   PCLO_HUMAN      STANDARD;      PRT;   5147 AA.
AC   Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragments).
GN   PCLO OR ACZ OR KIAA0559.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 4405-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9Y6V0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
```

```
CC                   VSP_003926, VSP_003927;
CC          Note=No experimental confirmation available;
CC      -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC          phospholipids. Calcium binds with low affinity but with high
CC          specificity and induces a large conformational change.
CC      -!- SIMILARITY: Contains 2 C2 domains.
CC      -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC      --------------------------------------------------------------------------
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      --------------------------------------------------------------------------
DR      EMBL; Y19188; CAB60727.1; -.
DR      EMBL; AC004903; AAD20936.1; -.
DR      EMBL; AC004886; AAD21789.1; -.
DR      EMBL; AB011131; BAA25485.1; -.
DR      EMBL; BC001304; AAH01304.1; -.
DR      EMBL; AC004082; AAB97937.1; -.
DR      PIR; T00634; T00634.
DR      HSSP; P04410; 1A25.
DR      Genew; HGNC:13406; PCLO.
DR      MIM; 604918; -.
DR      GO; GO:0005856; C:cytoskeleton; NAS.
DR      GO; GO:0045202; C:synaptic junction; ISS.
DR      GO; GO:0005509; F:calcium ion binding; ISS.
DR      GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR      GO; GO:0005522; F:profilin binding; ISS.
DR      GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR      GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR      InterPro; IPR000008; C2.
DR      InterPro; IPR001565; Synaptotagmin.
DR      PRINTS; PR00360; C2DOMAIN.
DR      PRINTS; PR00399; SYNAPTOTAGMN.
DR      SMART; SM00239; C2; 2.
DR      PROSITE; PS00499; C2_DOMAIN_1; 1.
DR      PROSITE; PS50004; C2_DOMAIN_2; 2.
KW      Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW      Repeat; Alternative splicing.
FT      NON_TER        1       1
FT      DOMAIN       400     465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                     P-A-K-P-Q-P-Q-Q-P-X.
FT      ZN_FING      499     523       C4-TYPE (POTENTIAL).
FT      ZN_FING      969     992       C4-TYPE (POTENTIAL).
FT      NON_CONS    1010    1011
FT      DOMAIN      2300    2325       POLY-PRO.
FT      DOMAIN      4391    4442       PDZ.
FT      DOMAIN      4544    4633       C2 DOMAIN 1.
FT      DOMAIN      5031    5121       C2 DOMAIN 2.
FT      VARSPLIC    4404    4404       S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                                     QTGKLMEG (in isoform 2).
FT                                     /FTId=VSP_003923.
FT      VARSPLIC    4534    4534       K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
FT                                     /FTId=VSP_003924.
FT      VARSPLIC    4576    4576       G -> GQVMVVQNAS (in isoform 2).
FT                                     /FTId=VSP_003925.
FT      VARSPLIC    4757    4761       TAHKS -> SKRRK (in isoform 2).
FT                                     /FTId=VSP_003926.
FT      VARSPLIC    4762    5147       Missing (in isoform 2).
FT                                     /FTId=VSP_003927.
SQ      SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

  Alignment Scores:
  Pred. No.:                  0.0044        Length:           5147
  Score:                      161.00        Matches:          212
  Percent Similarity:         31.35%        Conservative:     151
  Best Local Similarity:      18.31%        Mismatches:       437
  Query Match:                2.38%         Indels:           359
  DB:                         1             Gaps:             48
```

```
US-09-302-812-5 (1-3814) x PCLO_HUMAN (1-5147)

Qy         81 TGGAACTTCTGCGCCGACTGC---------------------CTCGGACTCCCGGAGCTT 119
              :::|||   |||   :::|||                          |||
Db        986 PheAsnThrCysThrGluCysLysAsnGlnValCysAsnLeuCysGlyPheAsnProThr 1005

Qy        120 CCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCC 179
              |||         |||  |||   |||   |||   :::          |||   :
Db       1006 ProHisLeuThrGluAsnCys--GlnThrGlnArgAlaIleSerGlyGlnLeuGlyAspI 1025

Qy        180 TCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGG------------------------- 214
              :::|||      ||||||   |||   :::
Db       1025 leArgLysMetProProAlaProSerGlyProLysAlaSerProMetProValProThrG 1045

Qy        215 --------GACCGCACAGAGGCAACGCCACCTCGTTT----------------------- 243
                      :::      |||   ||||||:::
Db       1045 luSerSerSerGlnLysThrAlaValProProGlnValLysLeuValLysLysGlnGluG 1065

Qy        244 -------------------------GTTTTCAAACAAAAGACTATTACTACTTGGATGG 277
                                       |||      :::|||      |||
Db       1065 lnGluValLysThrGluAlaGluLysValIleLeuGluLysValLysGluThrLeuSerM 1085

Qy        278 ATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTG 337
                  |||   |||         ::::::   |||:::      :::
Db       1085 etGluLysIleProProMetValThrThrAspGlnLysGlnGluGluSerLysLeuGluL 1105

Qy        338 ACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCA------------------- 378
                       |||::::::|||:::         |||
Db       1105 ysAspLysAlaSerAlaLeuGlnGluLysLysProLeuProGluGluLysLysLeuIleP 1125

Qy        379 --------CATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCAC 430
                      |||:::         |||      |||   |||      :::|||   |
Db       1125 roGluGluGluLysIleArgSerGluGluLysLysProLeuLeuGlu---GluLysLysP 1144

Qy        431 CCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGC 490
              |||||   :::      :::   |||   :::   :::|||::::::
Db       1144 roThrProGluAspLysLysLeuLeuProGluAlaLysThrSerAla------------P 1160

Qy        491 AGAATGAAGGGAAGCAT------------GCAGAACAGCTTTTGGCAAGTGAGCCTCCCG 538
                 |||   ||||||            ::::::      :::|||   |||
Db       1160 roGluGluGlnLysHisAspLeuLeuLysSerGlnValGlnIleAlaGluGluLysLeuG 1180

Qy        539 CGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTG 598
                 |||      ||||||   :::               |||:::   |||   |||
Db       1180 luGlyArgValAlaProLysThrValGlnGlu---------GlyLysGlnProGlnThrL 1197

Qy        599 ATGACCAC---------AGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTC 649
                               |||   |||         ::::::::   :::
Db       1197 ysMetGluGlyLeuProSerGlyThrProGlnSerLeuProLysGluAspAspLys---- 1215

Qy        650 TTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCA 709
                                 ||||||         |||
Db       1216 --------------------ThrThrLysThrIleLysGluGlnProGlnProProCysT 1229

Qy        710 ACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGG 769
                 |||         ::::::      |||      |||      |||:::
Db       1229 hrAlaLysProAspGlnGluLysGluAspAspLysSerAspThrSerSerSerGlnGln- 1248

Qy        770 AGGTGGATGTGCTACCAGAGAGTCCT------TTGTCAGATGTTGGTGCCGAGGACATTG 823
                                |||:::||||||      |||||||||
Db       1249 --------------ProLysSerProGlnGlyLeuSerAsp------------------- 1257

Qy        824 GAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGC 883
                 ||||||      :::|||                  |||   |||||||||||:::   |
Db       1258 --ThrGlyTyrSerSerAsp------------GlyIleSerSerSerLeuGlyGluIleP 1273

Qy        884 CT-----------------------------------------------------CCATTTG 892
              ||                                                      |||
Db       1273 roSerLeuIleProThrAspGluLysAspIleLeuLysGlyLeuLysLysAspSerPheS 1293

Qy        893 AGAAAGAAAGTGAGCCTGAGTCACCAATGGAT---------------------------- 924
                 :::||||||   |||   ||||||   |||
Db       1293 erGlnGluSerSerProSerSerProSerAspLeuAlaLysLeuGluSerThrValLeuS 1313
```

```
Qy    925 -----GTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTC 979
               ::::::          ::::::        |||   :::::::::::::::|||   |
Db   1313 erIleLeuGluAlaGlnAlaSerThrLeuAlaAspGluLysSerGluLysLysThrGlnP 1333

Qy    980 CA------GTCTTTGATGAGCAA------GATGATCGTTCCTCCCAAACAGCAAATAAA- 1026
          ||        |||       ||||||        |||         :::|||:::   ::::::
Db   1333 roHisGluValSerProGluGlnProLysAspGlnGluLysThrGlnSerLeuSerGluT 1353

Qy   1027 --------------CTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAAC 1072
                          :::|||        :::    :::||||:::       :::     ||||||
Db   1353 hrLeuGluIleThrIleSerGluGluGluIleLysGluSerGlnGluGlu---ArgLysA 1372

Qy   1073 GGTATTTGACTAAGGGAAGTGAAGTT---------------------------------- 1098
                          |||    |||:::
Db   1372 spThrPheLysLysAspSerGlnGlnAspIleProSerSerLysAspHisLysGluLysS 1392

Qy   1099 ---------------------------------------------AGATTGCATTTCCAAT 1114
                                                          :::
Db   1392 erGluPheValAspAspIleThrThrArgArgGluProTyrAspSerValGluGluSerS 1412

Qy   1115 TTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTT 1174
             |||     ||||||:::                                 |||::::||||
Db   1412 erGluSerGluAsnSerProValProGlnArgLysArgArgThrSerValGlySerSerS 1432

Qy   1175 CTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATT- 1233
          |||||             :::         |||:::       |||:::     :::       |||
Db   1432 erSerAspGluTyrLysGlnGluAspSerGlnGlySerGlyGluGluGluAspPheIleA 1452

Qy   1234 ----------------------------------------------------ACAGATCATT 1243
                                                                  |||    |
Db   1452 rgLysGlnIleIleGluMetSerAlaAspGluAspAlaSerGlySerGluAspAspGluP 1472

Qy   1244 TCATGAGA------------------ATTTCCAAGTCAGAGGACAGAAGAAAAGAACAAT 1285
          ||:::|||                         |||      :::|||     ::::::||||||:::
Db   1472 heIleArgAsnGlnLeuLysGluIleSerSerSerThrGluSerGlnLysLysGluGluT 1492

Qy   1286 GTGAA----------------------GTCAGACATCAAAGAACAGAAAGGAAGATTCCAA 1324
             :::                         :::|||:::|||        ||||||
Db   1492 hrLysGlyLysGlyLysIleThrAlaGlyLysHisArgArgLeuThrArgLysSerSerT 1512

Qy   1325 AATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAA 1384
                |||       :::                   |||                |||
Db   1512 hrSerIleAspGluAspAlaGlyArgArgHisSerTrpHisAspGluAspAspGluAlaP 1532

Qy   1385 TGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACA 1444
                :::     |||                                :::     |||
Db   1532 heAspGluSerProGluLeuLysTyrArgGluThrLysSerGlnGluSerGluGluLeuV 1552

Qy   1445 CAGTGACTGTTCGGGTAGACCTTCTGAGA------GCAGGAGAGGTTCCGAAACCTTTTC 1498
             ||||||               |||     |||                |||:::
Db   1552 alValThrGlyGlyGlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrIleA 1572

Qy   1499 CAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAA- 1557
                      |||                       |||           :::           |||:::
Db   1572 laAspLysTyrSerAlaGluSerSerGlnLysLysThrSerLeuTyrPheAspGluGluP 1592

Qy   1558 --------AACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGT 1609
                      :::        :::     |||        :::|||:::     |||
Db   1592 roGluLeuGluMetGluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlyS 1612

Qy   1610 GGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTG------- 1662
                   :::       :::                |||     |||      ::::::
Db   1612 erSerSerLeuHisAlaSerSerPheThrProGlyThrSerProThrSerValSerSerL 1632

Qy   1663 -----------AAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACT 1711
                        |||:::          :::                  ||||||
Db   1632 euAspGluAspSerAspSerSerProSerHisLysLysGlyGluSerLys---------- 1648

Qy   1712 TTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTAT 1771
                                                   ::::::     :::|||:::|||
Db   1649 -----------------------------------GlnGlnArgLysAlaArgHisArgP 1657

Qy   1772 ATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATT------- 1824
                   |||                                      ||||||     |||
Db   1657 roHisGlyProLeu---------------------------LeuProThrIleGluAspS 1668

Qy   1825 ------------------TGCACCCAGCCAATACCACTCCTGAAA---CAGAAGATGAATC 1864
                              :::                  |||||||||     |||:::
Db   1668 erSerGluGluGluGluLeuArgGluGluGluGluLeuLeuLysGluGlnGluLysGlnA 1688

Qy   1865 ATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCA 1924
             :::          |||    :::     :::|||
Db   1688 rgGluIleGluGlnGlnArgLysSerSerSer------------------------ 1699

Qy   1925 CATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACT 1984
                  ::::::::     |||     |||     |||
Db   1700 --------LysLysSerLysLysAspLysAspGluLeu--------------------- 1709

Qy   1985 TCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCT 2044
                                     |||:::     |||:::     |||:::     ||||||
Db   1710 -------------------ArgAlaGlnArgArgArgGluArgProLysThrProProS 1723

Qy   2045 GCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGA 2104
                          :::     :::          ||||||        :::             ||||||:
Db   1723 erAsnLeuSerProIleGluAspAlaSerProThrGluGluLeu---------ArgGlnA 1740

Qy   2105 GTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCA 2164
          ::     |||                                     :::      ||||||
Db   1740 laAlaGlu-------------------------------MetGluGluLeuHisArgS 1749

Qy   2165 CT------------TACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGG 2212
          ::                   |||    :::||||||     :::          |||
Db   1749 erSerCysSerGluTyrSerProSerIleGluSerAspProGluGly------------- 1764

Qy   2213 ATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAA 2272

Db   1764 ------------------------------------------------------------ 1764

Qy   2273 TCAGATTTTTAATCAATCCTGAATTGATTGTTTCA---------------CGGCTGTTCA 2317
                   |||     |||:::||||||     |||:::                      :::|||   |
Db   1765 -----PheGluIleSerProGluLysIleIleGluValGlnLysValTyrLysLeuProT 1783

Qy   2318 CTGAGGTG----------------------------------------CTGGATCACAATG 2338
          ||     |||                                          :::     ::::
Db   1783 hrAlaValSerLeuTyrSerProThrAspGluGlnSerIleMetGlnLysGluGlySerG 1803

Qy   2339 AGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTT 2398
          ::         :::     :::          |||     |||     |||                |
Db   1803 lnLysAlaLeuLysSerAlaGluGluMetTyrGluGluMetMetHisLysThrHisLysT 1823

Qy   2399 ATCGTTGGGCCCGAAGCCATGAAGATGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCA 2458
          ||                ::::::          :::|||:::|||
Db   1823 yr---------LysAlaPheProAlaAlaAsnGluArgAsp------------------ 1833

Qy   2459 CGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTG 2518
             |||:::          :::     |||:::                 ::::::   ::::::   |
Db   1834 --GluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAspTyrIleTyrG 1853

Qy   2519 AGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTT 2578
          ||     :::          :::          |||     |||                  |||
Db   1853 luSerLeu---------ValGluAspThrTyrAsnGly------------SerValAspG 1866

Qy   2579 CTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACG 2638
                  |||:::                   |||                |||    :::
Db   1866 lySerLeuLeuThrArgGlnGluGluGluAsnGlyPheMetGlnGlnLysGlyArgGluG 1886

Qy   2639 CTAGATTAAAAGCCTTAATACAGATC-------------------------------- 2664
            :::::::::          ||||||
Db   1886 lnLysIleArgLeuSerGluGlnIleTyrGluAspProMetGlnLysIleThrAspLeuG 1906

Qy   2665 --------------------------CTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTT 2698
                                       |||     :::          :::     |||:::|||
Db   1906 lnLysGluPheTyrGluLeuGluSerLeuHisSerValValProGlnGluAspIleValS 1926

Qy   2699 ATTTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTA 2758
```

```
                :::|||          |||         :::|||     :::   |||  :::|
Db       1926 erSerSerPheIleIleProGlu---SerHisGluIleValAspLeuGlyThrMetValT 1945

Qy       2759 CC-----------GAGAGGAAG---CTGGATGTTGGAAAAGTGTACAAGTTATTGCTTA 2803
               ||          |||||||||   ||||||           |||:::   |||::::
Db       1945 hrSerThrGluGluGluArgLysLeuLeuAspAlaAspAlaAlaTyrGluGluLeuMetL 1965

Qy       2804 GATACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATC 2863
              ::          :::                ||||||||                |
Db       1965 ysArgGlnGlnMetGlnLeu------------ThrProGlySerSerProThrGlnAlaP 1981

Qy       2864 CATTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGACATGCCA 2913
               ||                  ||||||:::   :::          ||||||
Db       1981 roIleGlyGluAspMetThrGluSerThrMetAspPheAspArgMetPro 1997

RESULT 9
NKCR_HUMAN
ID   NKCR_HUMAN     STANDARD;      PRT;  1462 AA.
AC   P30414;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE   related protein) (NK-TR protein).
GN   NKTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93133824; PubMed=8421688;
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS.
RA   Anderson S.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Component of a putative tumor-recognition complex.
CC       Involved in the function of NK cells.
CC   -!- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC       via its N-terminus.
CC   -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04288; AAA35734.2; -.
DR   EMBL; AF184110; AAD56402.1; -.
DR   PIR; A47328; A47328.
DR   HSSP; Q27450; 1A33.
DR   Genew; HGNC:7833; NKTR.
DR   MIM; 161565; -.
DR   GO; GO:0016018; F:cyclosporin A binding; TAS.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
FT   DOMAIN        1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN      219    240       ARG/LYS-RICH (BASIC).
FT   DOMAIN      421    457       ARG/LYS-RICH (BASIC).
FT   DOMAIN      970   1010       ARG/LYS-RICH (BASIC).
FT   DOMAIN      194    244       ARG/SER-RICH.
FT   DOMAIN      466    574       ARG/SER-RICH.
FT   DOMAIN      664    814       ARG/SER-RICH.
FT   DOMAIN     1311   1348       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE   1462 AA;  165676 MW;  D98A1147763EF527 CRC64;

Alignment Scores:
Pred. No.:                0.0032         Length:         1462
Score:                    160.50         Matches:        143
Percent Similarity:       33.47%         Conservative:   102
Best Local Similarity:    19.54%         Mismatches:     281
Query Match:              2.37%          Indels:         206
DB:                       1              Gaps:           30

US-09-302-812-5 (1-3814) x NKCR_HUMAN (1-1462)

Qy         85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
              |||::::::   ::::::|||   |||||||||
Db        537 ThrAlaSerLysSerSerSerHisSerArgSer--------------------------- 547

Qy        145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
                                                       |||   :::   |||
Db        548 ---------------------------------------ArgSerLysSerArgSerSer 554

Qy        205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
                :::::|||   ||||||   |||
Db        555 SerLysSerGly---HisArgLysArgAla------------------------------ 563

Qy        265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAAC 324
                             :::|||   |||:::         |||:::      ||||||
Db        564 ---------------SerLysSerProArgLysThrAlaSerGlnLeuSerGluAsnLys 578

Qy        325 AATACAAGAATTGACTCCATGATGAGTTCTGTG---CAGAAAGATAACTTTTACCCACAT 381
                 :::   :::   :::   :::::::::   |||   :::|||
Db        579 ProValLysThrGluProLeuArgAlaThrMetAlaGlnAsnGluAsnValValValGln 598

Qy        382 AAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCC--------- 432
                 |||      ||||||:::|||   :::   |||      ||||||
Db        599 ProVal---ValAlaGluAsnIleProValIleProLeuSerAspSerProProProSer 617

Qy        433 ---------------------------------------------------------ACA 435

Db        618 ArgTrpLysProGlyGlnLysProTrpLysProSerTyrGluArgIleGlnGluMetLys 637

Qy        436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
                 |||::::::      |||      |||   |||   |||:::         :::
Db        638 AlaLysThrThrHisLeuLeuProIleGlnSerThrTyrSerLeuAlaAsnIleLysGlu 657

Qy        496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
                 |||                                  :::   :::
Db        658 ThrGly------------------------------------SerSerSerSerTyrHis 665

Qy        556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACT------GATGACCACAGT 609
              |||:::      |||:::         ||||||   ::::::      |||   |||
Db        666 LysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSerAspArgSer 685

Qy        610 GACACAGATCATGAAGAAGACAGAGAC-----------------------AATCAGCAG 645
                          |||                                   ::::::
Db        686 SerGluSerSerProArgSerArgSerArgSerSerArgSerArgSerTyrSerArgSer 705

Qy        646 TTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGA 705
              :::                  ||||||::::::                   ::::::|||
Db        706 TyrThrArgSerArgSerLeuAlaSerSer---------------HisSerArgSerArg 720

Qy        706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765
              |||               |||   ||||||            |||   :::   |||   :::
Db        721 SerProSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArg 740

Qy        766 GAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGA 825
                                   |||   :::   :::   ::::::|||   |||
Db        741 --------------------SerSerSerTyrThrSerIleSerSerAspAsp---Gly 752

Qy        826 ACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCT 885
```

```
                 |||          :::    :::|||:::::::::|||:::
Db     753 ArgArgAlaLysArgArgLeuArgSerSerGlyLysLysAsnSerValSerHisLysLys 772

Qy     886 ---CCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAAC 942
                          |||            |||                   |||:::
Db     773 HisSerSerSerSerGluLysThrLeuHisSerLysTyrValLysGlyArgAspArgSer 792

Qy     943 AGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT 1002
           ||||||          ::::::        :::|||       :::          :::
Db     793 SerCysValArgLysTyrSerGluSerArgSerSerLeuAspTyrSerSerAspSerGlu 812

Qy    1003 CGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGAT 1062
           :::||||||      |||    :::        |||   |||   :::|||      |||
Db     813 GlnSerSerValGlnAlaThrGln------SerAlaGlnGluLysGluLysGlnGlyGln 830

Qy    1063 CTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGA 1122
           :::   :::   :::       :::
Db     831 MetGluArgThrHisAsnLysGln------------------------------------ 838

Qy    1123 GAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTT 1182
           |||   |||   |||                                   |||
Db     839 GluLysAsnArgGlyGlu---------------------------GluLysSerLysSer 849

Qy    1183 AATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCAT 1242
                 ||||||      ||||||:::          :::::::::   :::::||||||
Db     850 GluArgGluCysProHisSerLysLysArgThrLeuLysGluAsn---LeuSerAspHis 868

Qy    1243 TTCATGAGAATTTCCAAGTCAGAG---------------------------GACAGAAGA 1275
                    ||||||   :::                            |||   ||
Db     869 LeuArgAsnGlySerLysProLysArgLysAsnTyrAlaGlySerLysTrpAsp-SerGl 888

Qy    1276 AAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCA 1335
           |   |||:::   :::        :::||||||    |||    ||||||
Db     888 uSerAsnSerGluArg-AspValThrLysAsnSerLysAsnAspSerHis---------- 904

Qy    1336 CCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATG 1395
                   ||||||   ::::::       |||    :::         :::|||
Db     905 --------ProSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluSerGluV 922

Qy    1396 CCTCG------------------------------------------------------- 1400
             |||
Db     922 alSerGluIleHisIleLysValLysProThrThrLysSerSerThrAsnThrSerLeuP 942

Qy    1401 --------------GTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACA 1446
                             |||                        :::      |||
Db     942 roAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerThrSerAspSerGluG 962

Qy    1447 GTGACTGTTCG------------------GGTAGACCTTCTGAGAGCAGGAGAGGTTCC- 1487
                 ||||||                  |||:::|||    :::    :::    ||||||
Db     962 lySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHisLysHisGlySerL 982

Qy    1488 --------------------------GAAACCTTTTCCAACACATTACAAAGATTTGTGG 1521
                                      :::                |||  |||
Db     982 ysGluAsnLeuLysArgGluHisThrLysLysValLysGluLysLeu-LysGlyLysLys 1001

Qy    1522 GATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTAC---------CCTGTG 1572
           |||   ||||||   |||   |||         :::|||      :::         |||:::
Db    1002 AspLysLysHis---LysAlaProLysArgLysGlnAlaPheHisTrpGlnProProLeu 1020

Qy    1573 GAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTT 1632
           |||               |||                     ::::::   |||   :::
Db    1021 GluPheGlyGluGluGluGluGluGluIleAspAspLysGlnValThrGlnGluSerLys 1040

Qy    1633 CTCAACAAATTCACA---CGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTG 1689
                 |||   :::          :::   :::||||||   |||||||||
Db    1041 GluLysLysValSerGluAsnAsnGluThrIleLysAspAsnIleLeuLysThrGluLys 1060

Qy    1690 GCA------------TATTCTAAGAAATGGACTTTACAGCTTTGGTTGATTTCTGGGAT 1737
           :::             |||   |||   |||            :::             |||
Db    1061 SerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLeuAsp 1080

Qy    1738 AAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTG 1797
           :::      ::::::                     :::      :::   |||
Db    1081 GlnPheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrGluGlu 1100

Qy    1798 AAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAG 1857
              :::|||   ||||||   ||||||                              |||
Db    1101 AsnValAla---CysLeuGlnAsnIle-----------------------------GlnHis 1110

Qy    1858 ATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTC 1917
              :::      ||||||               |||    :::              |||
Db    1111 ValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMetGlu 1130

Qy    1918 TTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGT--------- 1968
              ||||||   |||   |||::::::      |||   |||   :::|||
Db    1131 IleCysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLeuGly 1149

Qy    1969 ---------------TACCCAGACATTAAC 1983
                          ||||||||||||
Db    1150 AsnAlaArgLeuAspThrProAspIleAsn 1159
```

```
RESULT 10
DSPP_RAT
ID   DSPP_RAT       STANDARD;      PRT;   687 AA.
AC   Q62598; P70578; Q9R057;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin sialophosphoprotein precursor [Contains: Dentin phosphoprotein
DE   (Dentin phosphophoryn) (DPP); Dentin sialoprotein (DSP)].
GN   DSPP OR RDSP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM DPP-2).
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20435277; PubMed=10978503;
RA   Ritchie H.H., Wang L.-H.;
RT   "The presence of multiple rat DSP-PP transcripts.";
RL   Biochim. Biophys. Acta 1493:27-32(2000).
RN   [2]
RP   SEQUENCE OF 1-387 FROM N.A., AND SEQUENCE OF 18-26.
RC   STRAIN=Sprague-Dawley; TISSUE=Odontoblast;
RX   MEDLINE=94148875; PubMed=8106414;
RA   Ritchie H.H., Hou H., Veis A., Butler W.T.;
RT   "Cloning and sequence determination of rat dentin sialoprotein, a
RT   novel dentin protein.";
RL   J. Biol. Chem. 269:3698-3702(1994).
RN   [3]
RP   SEQUENCE OF 421-687 FROM N.A. (ISOFORM DPP-1).
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=96355551; PubMed=8702961;
RA   Ritchie H.H., Wang L.-H.;
RT   "Sequence determination of an extremely acidic rat dentin
RT   phosphoprotein.";
RL   J. Biol. Chem. 271:21695-21698(1996).
RN   [4]
RP   SEQUENCE OF 1-19 FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=99333695; PubMed=10403786;
RA   Yamazaki H., Kunisada T., Miyamoto A., Tagaya H., Hayashi S.-I.;
RT   "Tooth-specific expression conferred by the regulatory sequences of
RT   rat dentin sialoprotein gene in transgenic mice.";
RL   Biochem. Biophys. Res. Commun. 260:433-440(1999).
RN   [5]
RP   SEQUENCE OF 29-33; 70-79; 93-109; 136-148; 162-188; 266-308; 398-423
RP   AND 426-438, AND PHOSPHORYLATION OF SER-292 AND SER-298.
RX   MEDLINE=21125612; PubMed=11042175;
RA   Qin C., Cook R.G., Orkiszewski R.S., Butler W.T.;
RT   "Identification and characterization of the carboxyl-terminal region
RT   of rat dentin sialoprotein.";
RL   J. Biol. Chem. 276:904-909(2001).
RN   [6]
RP   TISSUE SPECIFICITY.
```

```
RX   MEDLINE=98055479; PubMed=9395101;
RA   Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
RA   Hotton D., Papagerakis P., Berdal A., Butler W.T.;
RT   "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
RT   expression in odontoblasts and transient expression in
RT   pre-ameloblasts.";
RL   Eur. J. Oral Sci. 105:405-413(1997).
CC   -!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
CC       may bind high amount of calcium and facilitate initial
CC       mineralization of dentin matrix collagen as well as regulate the
CC       size and shape of the crystals.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPP-1; Synonyms=PP240;
CC         IsoId=Q62598-1; Sequence=Displayed;
CC       Name=DPP-2; Synonyms=PP171;
CC         IsoId=Q62598-2; Sequence=VSP_003855;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Specifically expressed in teeth, mainly in
CC       odontoblasts and transiently in pre-ameloblasts.
CC   -!- PTM: DSP is glycosylated.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 380.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF247187; AAK96895.1; -.
DR   EMBL; U02074; AAA18932.1; ALT_FRAME.
DR   EMBL; U63111; AAC52774.1; -.
DR   EMBL; AF114987; AAD48588.1; ALT_TERM.
KW   Biomineralization; Extracellular matrix; Signal; Glycoprotein;
KW   Phosphorylation; Sialic acid; Alternative splicing.
FT   SIGNAL        1     17
FT   CHAIN        18    687       DENTIN SIALOPHOSPHOPROTEIN.
FT   CHAIN        18    447       DENTIN SIALOPROTEIN.
FT   CHAIN       448    687       DENTIN PHOSPHOPROTEIN.
FT   MOD_RES      57     57       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     226    226       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     253    253       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     278    278       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     292    292       PHOSPHORYLATION (BY CK2).
FT   MOD_RES     298    298       PHOSPHORYLATION (BY CK1).
FT   MOD_RES     315    315       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     319    319       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     329    329       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     337    337       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     345    345       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     366    366       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   CARBOHYD     55     55       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    189    189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    567    635       Missing (in isoform DPP-2).
FT                                /FTId=VSP_003855.
FT   CONFLICT     74     74       N -> D (IN REF. 5; AA SEQUENCE).
FT   CONFLICT    564    564       S -> T (IN REF. 3).
SQ   SEQUENCE   687 AA;  70179 MW;  9A845EED6AA31B63 CRC64;

Alignment Scores:
Pred. No.:                 0.0032        Length:          687
Score:                     159.00        Matches:         85
Percent Similarity:        34.84%        Conservative:    61
Best Local Similarity:     20.29%        Mismatches:      195
Query Match:               2.35%         Indels:          78
```

```
DB:                    1             Gaps:         13

US-09-302-812-5 (1-3814) x DSPP_RAT (1-687)

Qy         73 GGCGCCGCTGGAACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAG 132
              |||      ||||||:::        :::                  |||    |||
Db        309 GlyGlyAspAsnThrSerSerSerGluGluThrGlyIleGluGluGlyAspGlyThrGln 328

Qy        133 AGGCGTGTTCTCGACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCA 192
                       :::       :::|||   :::      :::         :::
Db        329 ThrThrGlnAspAsnGlnAsnLeuSerProThrGluGlyGlyIleIleSerGlnAlaGlu 348

Qy        193 GCCTGCGTCTCGGGGCGGGCG-----------------------GGACCGCACAGAGGC 228
              ||||||    ||||||::::::                     |||         |||
Db        349 AlaCysProSerGlyGlnSerGlnAsnGlnGlyLeuGluThrGluGlySerSerThrGly 368

Qy        229 AACGCCACCTCGTTTGTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGA 288
              |||   :::|||          :::         ::::::        |||:::   |||
Db        369 AsnLysSerSerIleThrLysGluSerGlyLysLeuSerGlySerLysAspSerAsnGly 388

Qy        289 ------------------------------CCCAAGACAGCTGAATCAGAA------ 309
                                            ||||||      ||||||:::
Db        389 HisHisGlyMetGluLeuAspLysArgAsnSerProLysGlnGlyGluSerAspLysPro 408

Qy        310 ----------------AGTAAAGAAAACAACAAT---------ACAAGAATTGACTCC 342
                              |||      :::||||||       :::||||||   |||
Db        409 GlnGlyAlaAlaGluLysSerAspThrHisAsnAsnMetGlyHisSerArgIleGlySer 428

Qy        343 ATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAAT 402
                 :::|||          |||:::   |||               |||   ::::::
Db        429 SerSerAsnSerAspGlyHisAspSer---TyrAspPheAspAspGluSerMetGlnGly 447

Qy        403 GTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAACCAA 462
                       :::    |||:::|||                  |||         |||
Db        448 AspAspProAsnSerSerAspGluSerAsnGlySerAspGlySerAspAspAlaAsnSer 467

Qy        463 CAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTG 522
              :::                                 |||   |||      :::
Db        468 GluSerAlaIle-----------------GluAsnGlyAsnHisGlyAspAlaSerTyr 481

Qy        523 GCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGT 582
                 |||:::      :::                         |||:::::::::   |||
Db        482 ThrSerAspGluSerSerAspAsnGlySerAsp---------SerAspSerHisAlaGly 498

Qy        583 CAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAG 642
              :::         :::||||||   |||||||||||         :::         :::
Db        499 GluAspAspSerSerAspAspThrSerAspThrAspAspSerAspSerAsnGlyAspAsp 518

Qy        643 CAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCC 702
                       :::      |||                                |||
Db        519 AspSerGluSerLysAspLys--------------------------AspGluSerAsp 529

Qy        703 AGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAA 762
                 ||||||                         :::|||      |||   :::
Db        530 AsnSerAsnHisAspAsnAspSerAspSerGluSerLysSerAspSerSerAspSerAsp 549

Qy        763 CAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATT 822
                 :::      |||         |||      ||||||      :::|||
Db        550 SerAspSerSerAspSerSerAspSerSerAspSerSerAsp---SerSerGluThrSer 568

Qy        823 GGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCG 882
                 :::         :::|||         :::         |||         ||||||
Db        569 AspSerSerAspSerSerAspThrSerAspSerSerAspSerSerAspSerSerAspSer 588

Qy        883 CCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAAC 942
                          |||:::                     |||   |||:::|||
Db        589 SerAsnSerSerAspThrSerAspSerSerAspSerSerAspGlyAspSerSerAspGly 608

Qy        943 AGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT 1002
                    ||||||::::::   :::         |||                  |||
Db        609 AspSerSerAspSerAspSerSerAspSerAspSer---------SerAsnSerSerAsp 625

Qy       1003 CGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGAT 1062
```

```
               |||       ::::::       |||||       :::    ::::::|||   |||
Db        626 SerAspSerSerAspSerSerAspSerSerSerSerAspSerSerAspSerAspSerAsp 645

Qy       1063 CTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGA 1122
                 :::
Db        646 SerLys------------------------------------------------------ 647

Qy       1123 GAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTT 1182
              ::::::   :::   :::||||||   |||:::|||   |||||||||   :::|||
Db        648 AspSerThrSerAspSerSerAspAspAsnSerLys---SerGlyAsnGlyAsnSer--- 665

Qy       1183 AATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGAT 1239
                                 |||                  ||||||:::   :::|||
Db        666 ---------------AspSerAspSerAspSerAspSerAspSerGluGlySerAsp 679

RESULT 11
PCLO_RAT
ID   PCLO_RAT       STANDARD;      PRT;  5085 AA.
AC   Q9JKS6; Q9JLT1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN   PCLO.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX   MEDLINE=20170257; PubMed=10707984;
RA   Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA   Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT   "Piccolo, a presynaptic zinc finger protein structurally related to
RT   bassoon.";
RL   Neuron 25:203-214(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP   VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP   ALA-4694.
RX   MEDLINE=21181819; PubMed=11285225;
RA   Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT   "An unusual C(2)-domain in the active-zone protein piccolo:
RT   implications for Ca(2+) regulation of neurotransmitter release.";
RL   EMBO J. 20:1605-1619(2001).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC       junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9JKS6-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF138789; AAF07822.2; -.
DR   EMBL; AF227534; AAF63196.1; -.
DR   HSSP; P04410; 1A25.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; IDA.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      372    491       12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     523    547       C4-TYPE (POTENTIAL).
FT   ZN_FING    1010   1033       C4-TYPE (POTENTIAL).
FT   DOMAIN     2351   2362       POLY-PRO.
FT   DOMAIN     4442   4536       PDZ.
FT   DOMAIN     4653   4752       C2 DOMAIN 1.
FT   DOMAIN     4968   5059       C2 DOMAIN 2.
FT   VARSPLIC   4876   4880       TKPTN -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003930.
FT   VARSPLIC   4881   5085       Missing (in isoform 2).
FT                                /FTId=VSP_003931.
FT   MUTAGEN    4668   4668       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                BINDING ACTIVITY.
FT   MUTAGEN    4674   4674       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                BINDING ACTIVITY.
FT   MUTAGEN    4688   4688       V->S: SMALL INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4688   4689       VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4689   4689       M->S: INCREASED AFFINITY FOR CALCIUM.
FT   MUTAGEN    4690   4691       VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4692   4693       QN->AA: MODERATE INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4694   4694       A->S: NO EFFECT ON CALCIUM-BINDING
FT                                ACTIVITY.
SQ   SEQUENCE   5085 AA;  552702 MW;  5A1BB543201A7450 CRC64;

Alignment Scores:
Pred. No.:               0.00709        Length:        5085
Score:                   158.00         Matches:       147
Percent Similarity:      34.11%         Conservative:  101
Best Local Similarity:   20.22%         Mismatches:    272
Query Match:             2.33%          Indels:        208
DB:                      1              Gaps:          31

US-09-302-812-5 (1-3814) x PCLO_RAT (1-5085)

Qy         40 CCCGGCTGGGAGCCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACTTCTGCGCCGACT 99
                |||||       |||   |||        :::    |||         |||
Db       1249 ProGlyAlaThrProLeuThrLeuProLysAlaGlyGluLys-GluArgAlaValAlaGl 1268

Qy        100 GCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACG-- 157
                |||:::       |||:::|||      |||:::|||      |||         :::|||
Db       1268 nProGlnAlaGluGlySerSerLysAspGlyGlnGlyGluArgSerLysGluLysThrGl 1288
```

```
Qy    158 -------------CTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCT-------- 196
                       :::|||   |||   :::|||:::   |||
Db   1288 uLysGluGluAspLysSerAspThrSerSerSerGlnGlnProLysSerProGlnGlyLe 1308

Qy    197 ----------------GCGTCTCGGGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCG 240
                          :::|||   |||                       |||
Db   1308 uSerAspThrGlyTyrSerSerAspGlyIleSerGlySerLeuGlyGluIleProSerLe 1328

Qy    241 TTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCC----AAGAC 296
           :::   |||:::::::::   ||||||  |||  :::   |||||
Db   1328 uIleProSerAspGluLysAspLeuLeuLysGly---LeuLysLysAspSerPheSerGl 1347

Qy    297 AGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGT 356
             |||   |||       ::::::   :::::::::|||   :::   |||::
Db   1347 nGluSerSerProSerSerProSerAspLeuAlaLysLeuGluSerThrValLeuSerIl 1367

Qy    357 G---------CAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAAT----- 402
          :            |||                     |||  |||||
Db   1367 eLeuGluAlaGlnAlaSerThrLeuValGlyGluLysAlaGluLysLysThrGlnProGl 1387

Qy    403 -------GTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTATTT 455
                    |||:::         |||:::   |||:::  :::|||:::  ||
Db   1387 nLysIleSerProGluLysProGlnAspGlnGlnLysThrGlnThrAlaSerGluThrLe 1407

Qy    456 G--------------------------------AACCAACAGCAGACTGCGAGTGTGTG 482
          |                                :::|||:::::::     |||
Db   1407 uAspIleThrIleSerGluGluGluIleLysGluSerGlnGluLysLysValSerProLy 1427

Qy    483 CAAGTGGCAGAATGAAGGG---------------AAGCATGCAGAACAG----------- 516
           |||          :::|||           :::|||  |||:::
Db   1427 sLysAspSerGluGlnGlyPheProSerArgLysGluHisLysGluLysProGluLeuVa 1447

Qy    517 -------------------------------------CTTTTGGCAAGTGAGCCTCCCGC 539
                                               :::       |||
Db   1447 lAspAspLeuSerProArgArgAlaSerTyrAspSerValGluAspSerSerGluSerGl 1467

Qy    540 GGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGA 599
               :::|||:::  ::::::           :::||||||  |||      :::||
Db   1467 uAsnSerProValValArgArgLysArgArgThrSerIleGlySerSer---SerSerAs 1486

Qy    600 TGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCT-- 657
          |::::::           |||    :::          :::   :::   |||:::
Db   1486 pGluTyrLysGlnGluAspSerGlnGlySerGlyGluGluGluAspPheIleArgLysGl 1506

Qy    658 ----ATAAAACTTGCAAATACAAAGCCAACA----------------------------- 684
              |||:::::::::       :::
Db   1506 nIleIleGluMetSerAlaAspGluAspAlaSerGlySerGluAspGluGluPheIleAr 1526

Qy    685 -------------------------GTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTG 719
                                   ||||||:::   |||   |||        |||
Db   1526 gSerGlnLeuLysGluIleSerGlyValGlyGluSerGlnLysArgGluGluAlaLysGl 1546

Qy    720 CAGTGGA--------------------------------------TCTCGCCAGTCTGT 740
              |||                                      |||      |||
Db   1546 yLysGlyLysGlyValAlaGlyLysHisArgArgLeuThrArgLysSerSerThrSerPh 1566

Qy    741 GAAAGACTGTACAGGCTGTCAA---------------------CAGGAGGAGGTGGATGT 779
              |||      |||    :::                        |||      |||
Db   1566 eAspAspAspAlaGlyArgArgHisSerTrpHisAspGluAspAspGluThrPheAspGl 1586

Qy    780 GCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGAC------------------AT 821
             ||||||         :::      :::::::::
Db   1586 uSerProGluLeuLysPheArgGluThrLysSerGlnGluSerGluGluLeuValValAl 1606

Qy    822 TGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGAT-- 879
            |||  |||            |||            |||::::::  |||
Db   1606 aGlyGlyGlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrIleAlaAspLy 1626

Qy    880 ---------------------------------------------TCGCCTCCATT 890
                                                           |||
Db   1626 sTyrSerSerGluSerSerGlnLysLysThrIleLeuTyrPheAspGluGluProGluLe 1646

Qy    891 TGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGA------AACAG 944
           |||   |||||            :::||||||  |||                :::||
Db   1646 uGluMetGluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlySerSerSe 1666

Qy    945 TTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC------------TTTGATGA 992
          |          |||                  |||||||||            |||||
Db   1666 rLeuHisAlaSerSerPheThrProGlyThrSerProThrSerValSerSerLeuAspGl 1686

Qy    993 GCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGC 1052
          |         |||    ||||||    :::         |||       |||    |||    |||:::||
Db   1686 uAspSerAsp---SerSerProSerHisLysLysGlyGluSerLysGlnGlnArgLysAl 1705

Qy   1053 TGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCA 1112
          |                       |||:::   :::   |||      :::
Db   1705 a-----------------ArgHisArgSerHisGlyProLeuLeu-------------- 1714

Qy   1113 ATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTC 1172
                                                          |||:::
Db   1715 ------------------------------------------ProThrIleGluAs 1719

Qy   1173 TTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAA-- 1230
           ||||||           |||   |||         ::::::       |||        |||
Db   1719 pSerSerGluGluGluGluLeuArgGluGluGluGluLeuLeuLysGluGlnGluLysGl 1739

Qy   1231 ----ATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATG 1286
                :::          |||   |||   |||:::       :::|||::::::
Db   1739 nArgGluLeuGluGlnGlnGlnArgLysSerSerSerLysLysSerLysLysAspLysAs 1759

Qy   1287 TGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCC 1346
           |||:::|||   ||||||         |||:::   ||||||          |||   ||||||
Db   1759 pGluLeuArgAlaGlnArg---ArgArgGluArgProLysThrProProSerAsnLeuSe 1778

Qy   1347 TCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGG----- 1401
           |||    :::             :::|||   ||||||:::|||:::
Db   1778 rProIleGluAsp------AlaSerProThrGluGluLeuArgGlnAlaAlaGluMetGl 1796

Qy   1402 ------------------TGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCA 1442
                            |||          :::    ||||||:::
Db   1796 uGluLeuHisArgSerSerCysSerGluTyrSerProSerIleGluSerAsp-------- 1813

Qy   1443 CACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAC 1502
                                                    |||:::    |||
Db   1814 ----------------------------------------ProGluGlyPheGluIl 1819

Qy   1503 ACATTACAAAGATTTGTGGGAT---AACAAACATGTGAAAATGCCTTGTTCGGAACAAAA 1559
              :::   :::    :::                |||    |||:::|||    :::
Db   1819 eSerProGluLysIleIleGluValGlnLysValTyrLysLeuProAlaAlaValSerLe 1839

Qy   1560 CTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCAT 1619
                 |||    :::::::::::        :::        |||                 ::
Db   1839 uTyrSerProThrAspGluGlnSerValMetGlnLysGluGly---------------Va 1854

Qy   1620 TCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAA 1679
          :|||   |||                                    ||||||    |||
Db   1854 lGlnLysAla-----------------------------LeuLysSerAlaGluGluMe 1864

Qy   1680 ATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTG------GTTGATTTCTG 1733
           |||             |||     :::     |||             :::
Db   1864 tTyrGluGluMetMetGlnLysProHisLysTyrLysAlaPheProAlaAlaAsnGluAr 1884

Qy   1734 GGATAAGGTACTTGAAGAAGCAGAG---------------------GCCCAACATTTATA 1772
           |||:::|||   |||:::                              ::::::||
Db   1884 gAspGluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAspTyrIleTy 1904

Qy   1773 TCAGTCCATTTTACCTGAC 1791
          |:::|||::::::   |||
Db   1904 rGluSerLeuValGluAsp 1910
```

RESULT 12
ENL_HUMAN
ID ENL_HUMAN STANDARD; PRT; 559 AA.
AC Q03111;
DT 01-FEB-1994 (Rel. 28, Created)

```
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   ENL protein.
GN   MLLT1 OR ENL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93046667; PubMed=1423624;
RA   Tkachuk D.C., Kohler S., Cleary M.L.;
RT   "Involvement of a homolog of Drosophila trithorax by 11q23
RT   chromosomal translocations in acute leukemias.";
RL   Cell 71:691-700(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94362221; PubMed=8080983;
RA   Rubnitz J.E., Morrissey J., Savage P.A., Cleary M.L.;
RT   "ENL, the gene fused with HRX in t(11;19) leukemias, encodes a
RT   nuclear protein with transcriptional activation potential in lymphoid
RT   and myeloid cells.";
RL   Blood 84:1747-1752(1994).
CC   -!- FUNCTION: Capable of activating transcription from synthetic
CC       reporter genes in both lymphoid and myeloid cells.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Involved in acute leukemias through a chromosomal
CC       translocation t(11;19)(q23;p13.3) that involves MLL/HRX and MLLT1.
CC       The result is a rogue activator protein.
CC   -!- SIMILARITY: TO HUMAN MLLT3/AF9.
CC   -!- SIMILARITY: TO YEAST TRANSCRIPTION INITIATION FACTOR TFIIF SMALL
CC       SUBUNIT (TFG3/ANC1).
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/ENL.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04285; AAA58457.1; -.
DR   PIR; B44265; B44265.
DR   Genew; HGNC:7134; MLLT1.
DR   MIM; 159556; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR   InterPro; IPR005033; YEATS.
DR   Pfam; PF03366; YEATS; 1.
KW   Transcription regulation; Activator; Nuclear protein;
KW   Chromosomal translocation; Proto-oncogene.
FT   DOMAIN      268    280       POLY-PRO.
FT   DOMAIN      319    325       POLY-SER.
FT   DOMAIN      374    389       POLY-SER.
FT   DOMAIN      457    460       POLY-PRO.
FT   DOMAIN      463    466       POLY-PRO.
SQ   SEQUENCE   559 AA;  62082 MW;  A6836722EC84D004 CRC64;

Alignment Scores:
Pred. No.:               0.00381        Length:         559
Score:                   157.50         Matches:        110
Percent Similarity:      34.60%         Conservative:   63
Best Local Similarity:   22.00%         Mismatches:     205
Query Match:             2.33%          Indels:         122
DB:                      1              Gaps:           20

US-09-302-812-5 (1-3814) x ENL_HUMAN (1-559)

Qy       52 CCCTGCACGAAAGCGCGCTGG---------GGCGCCGCTGGAACTTCTGCG---CCGACT 99

Db      127 ProThrThrGluPheArgTyrLysLeuLeuArgAlaGlyGlyValMetValMetProGlu 146

Qy      100 GCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGACCCCAAGGACGCT 159

Db      147 GlyAlaAspThrValSerArgProSer-----------------------ProAspTyr 158

Qy      160 CCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGGGCGGGCGGGACCG 219

Db      159 ProMetLeuProThrIleProLeuSer--------------------------------- 167

Qy      220 CACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGAT 279

Db      168 -----------------AlaPheSerAspProLysLysThr------------LysPro 177

Qy      280 ACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGAC 339

Db      178 SerHisGlySerLysAspAlaAsnLysGluSerSerLysThrSerLys----------- 193

Qy      340 TCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAA 399

Db      194 -----------------------------------ProHisLysValThrLys---Glu 200

Qy      400 AATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAAC 459

Db      201 HisArgGluArgProArgLysAsp---SerGluSerLysSerSerSerLysGluLeuGlu 219

Qy      460 CAACAGCAGACTGCGAGT---------GTGTGCAAGTGGCAGAATGAAGGGAAGCATGCA 510

Db      220 ArgGluGlnAlaLysSerSerLysAspThrSerArgLysLeuGlyGluGlyArgLeuPro 239

Qy      511 GAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAG----------- 558

Db      240 LysGluGluLysAlaProProProLysAlaAlaPheLysGluProLysMetAlaLeuLys 259

Qy      559 ------------CAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCAC 606

Db      260 GluThrLysLeuGluSerThrSerProAsnProGlyProProProProProPro------ 277

Qy      607 AGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTT 666

Db      278 ---------------------------------------------ProProProArgAla 282

Qy      667 GCAAATACAAAGCCAACAGTAGGAGATGGG---------------------CAGGCCAGA 705

Db      283 SerSerLysArgProAlaThrAlaAspSerProLysProSerAlaLysLysGlnLysLys 302

Qy      706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765

Db      303 SerSerSerLysGlySerArgSerAlaProGlyThrSerProArgThrSerSerSerSer 322

Qy      766 GAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATT--- 822

Db      323 SerPheSerAspLysLysProAlaLysAspLysSerSerThrArgGlyGluLysValLys 342

Qy      823 ---GGAACTGGACCAAAAAATGACAACAAA---TTGACTGGACAAGAAAGCAGCCTAGGT 876

Db      343 AlaGluSerGluProArgGluAlaLysLysAlaLeuGluValGluGluSerAsnSerGlu 362

Qy      877 GATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGACAACTCG 936

Db      363 AspGluAlaSerPheLysSerGluSerAlaGlnSerSerProSerAsnSerSerSerSer 382

Qy      937 AGAAACAGTTGTCAAGATTCAGAA------------GCAGATGAAGAAACAAGTCCAGTC 984

Db      383 SerAspSerSerSerAspSerAspPheGluProSerGlnAsnHisSerGlnGlyProLeu 402

Qy      985 TTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCA 1044

Db      403 ArgSerMetValGluAspLeuGlnSerGluGluSerAspGluAspAspSerSerSerGly 422

Qy     1045 AGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTG 1104
```

```
Db        423 GluGluAlaAlaGlyLys---------------ThrAsnProGlyArgAspSerArgLeu 437

Qy       1105 CATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTA--------------- 1149
                 |||               :::|||:::       |||   |||
Db        438 SerPheSerAspSerGluSerAspAsnSerAlaAspSerSerLeuProSerArgGluPro 457

Qy       1150 ---------------------------AATGCCAAGCCATCTGGAAACTCTTCTAGCCTT 1182
                                         |||:::|||  ||||||     |||
Db        458 ProProProGlnLysProProProProAsnSerLysValSerGlyArgArgSerPro--- 476

Qy       1183 AATGTAGAGTGTAGAAGTTCCAAG---------CAGCATGGAAAAAGGGATTCTAAAATT 1233
                       |||      :::          :::  |||    |||
Db        477 ---GluSerCysSerLysProGluLysIleLeuLysLysGlyThrTyrAspLysAlaTyr 495

Qy       1234 ACAGATCATTTCATG------------AGAATTTCCAAGTCAGAGGACAGAAGAAAAGAA 1281
              ||||||      :::            |||:::            :::|||
Db        496 ThrAspGluLeuValGluLeuHisArgArgLeuMetAlaLeuArgGluArgAsnValLeu 515

Qy       1282 CAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAAC 1341
              |||      ||| ::::::  |||:::  |||   |||       |||:::
Db        516 GlnGlnIleVal-AsnLeuIleGluGluThrGlyHisPheAsnValThrAsnThrThrPh 535

Qy       1342 CTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTC 1399
                                |||      |||       :::||||||    :::   ||||||
Db        535 eAspPheAspLeuPheSerLeuAspGluThrThrValArgLysLeuGlnSerCysLeu 554
```

```
RESULT 13
DSPP_MOUSE
ID   DSPP_MOUSE     STANDARD;      PRT;   934 AA.
AC   P97399; O70567;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP-
DE   3) [Contains: Dentin phosphoprotein (Dentin phosphophoryn) (DPP);
DE   Dentin sialoprotein (DSP)].
GN   DSPP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Molar;
RX   MEDLINE=97150835; PubMed=8995371;
RA   MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;
RT   "Dentin phosphoprotein and dentin sialoprotein are cleavage products
RT   expressed from a single transcript coded by a gene on human chromosome
RT   4. Dentin phosphoprotein DNA sequence determination.";
RL   J. Biol. Chem. 272:835-842(1997).
RN   [2]
RP   SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RX   MEDLINE=98211969; PubMed=9545272;
RA   Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,
RA   D'Souza R.N., Kozak C.A., MacDougall M.;
RT   "Genomic organization, chromosomal mapping, and promoter analysis of
RT   the mouse dentin sialophosphoprotein (Dspp) gene, which codes for both
RT   dentin sialoprotein and dentin phosphoprotein.";
RL   J. Biol. Chem. 273:9457-9464(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RA   Sfeir C., Butler S., Lin E., George A., Veis A.;
RT   "From mouse to zebrafish-dentin matrix proteins genomic
RT   characterization.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=98055479; PubMed=9395101;
RA   Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
RA   Hotton D., Papagerakis P., Berdal A., Butler W.T.;
RT   "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
RT   expression in odontoblasts and transient expression in
RT   pre-ameloblasts.";
RL   Eur. J. Oral Sci. 105:405-413(1997).
RN   [5]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=21096982; PubMed=11175790;
RA   Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,
RA   Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,
RA   Zhao G., Kong X.;
RT   "Dentinogenesis imperfecta 1 with or without progressive hearing loss
RT   is associated with distinct mutations in DSPP.";
RL   Nat. Genet. 27:201-204(2001).
CC   -!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
CC       may bind high amount of calcium and facilitate initial
CC       mineralization of dentin matrix collagen as well as regulate the
CC       size and shape of the crystals.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and
CC       transiently in pre-ameloblasts. Found in the inner ear.
CC   -!- PTM: DSP is glycosylated.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U67916; AAC12787.1; -.
DR   EMBL; AJ002141; CAA05208.1; ALT_SEQ.
DR   EMBL; AF135799; AAD42781.1; ALT_SEQ.
DR   MGD; MGI:109172; Dspp.
KW   Biomineralization; Extracellular matrix; Signal; Glycoprotein;
KW   Phosphorylation; Sialic acid.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    934       DENTIN SIALOPHOSPHOPROTEIN.
FT   CHAIN        18    451       DENTIN SIALOPROTEIN.
FT   CHAIN       452    934       DENTIN PHOSPHOPROTEIN.
FT   DOMAIN      419    934       SER/ASP-RICH.
FT   SITE        479    481       CELL ATTACHMENT SITE (POTENTIAL).
FT   MOD_RES     227    227       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     254    254       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     279    279       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     293    293       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES     299    299       PHOSPHORYLATION (BY CK1) (BY SIMILARITY).
FT   MOD_RES     314    314       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     336    336       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     349    349       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    313    313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    373    373       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   934 AA;  93901 MW;  A618789D8A57249A CRC64;
```

```
Alignment Scores:
Pred. No.:              0.00448       Length:        934
Score:                  157.50        Matches:       94
Percent Similarity:     34.13%        Conservative:  64
Best Local Similarity:  20.30%        Mismatches:    172
Query Match:            2.33%         Indels:        133
DB:                     1             Gaps:          18

US-09-302-812-5 (1-3814) x DSPP_MOUSE (1-934)

Qy         19 GGTCCCAGCATGAGTGCGGGCCCCGGCTGGGAG------CCCTGCACGAAAGCGCGCTGG 72
              |||               ||||||   |||   |||         ||||||   :::
Db        257 GlyGluGlyAlaGluAlaGlyAspGlyArgGluSerHisAspGlyThrLysGlyGlnGly 276

Qy         73 GGCGCCGCT------------------------GGAACTTCTGCGCCGACTGCCTCGGAC 108
              |||   :::                        |||   |||:::   :::      |||
```

```
Db       277 GlyGlnSerHisGlyGlyAsnThrAspHisArgGlyGlnSerSerValSerThrGluAsp 296

Qy       109 TCCCGGAGC---------------TTCCCTGGC-------------------------- 126
                   |||               ||||||
Db       297 AspAspSerLysGluGlnGluGlyPheProAsnGlyHisAsnGlyAspAsnSerSerGlu 316

Qy       127 ---------------------------------------AGGCAGAGGCGTGTTCTCGAC 147
                                                    :::   :::       |||
Db       317 GluAsnGlyValGluGluGlyAspSerThrGlnAlaThrGlnAspLysGluLysLeuSer 336

Qy       148 CCCAAGGACGCT---CCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
             ||||||||             :::      :::         |||   ||||||   |||
Db       337 ProLysAspThrArgAspAlaGluGlyGlyIleIleSerGlnSerGluAlaCysProSer 356

Qy       205 GGGCGGGCG------------------------GGACCGCACAGAGGCAACGCCACCTCG 240
             |||::::::                        ||||||::::::||||||
Db       357 GlyLysSerGlnAspGlnGlyIleGluThrGluGlyProAsnLysGlyAsn--------- 373

Qy       241 TTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGATACTAAAGGACCCAAGACAGCT 300
                            |||:::|||  |||                |||      :::
Db       374 ---------------LysSerIleIleThr------LysGluSerGlyLysLeuSerGly 386

Qy       301 GAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAG 360
                   :::|||      :::            :::|||      |||         |||
Db       387 SerLysAspSerAsnGlyHisGlnGlyValGluLeuAspLysArgAsnSerProLysGln 406

Qy       361 AAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTT 420
                ::::::      |||
Db       407 GlyGluSerAspLysPro------------------------------------------ 412

Qy       421 GATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTG 480
                :::         |||||||||:::   :::   |||
Db       413 ---GlnGlyThrAlaGluLysSerAlaAlaHisSerAsn--------------------- 424

Qy       481 TGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCG 540

Db       424 ------------------------------------------------------------ 424

Qy       541 GGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGAT 600
                                   |||  ::::::   ||||||   |||   ::::::|||
Db       425 ---------------------LeuGlyHisSerArgIleGlySerSerSerAsnSerAsp 437

Qy       601 GACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATA 660
                |||            :::|||   |||   ::::::   |||            |||
Db       438 GlyHisAsp------SerTyrGluPheAspAspGluSerMetGlnGlyAspAspPro--- 454

Qy       661 AAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGC 720
                      ::::::      :::   |||      ::::::   :::|||      :::
Db       455 ------LysSerSerAspGluSerAsnGlySerAspGluSerAspThrAsnSerGluSer 472

Qy       721 AGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTG 780
             :::            |||      |||   :::            |||      |||
Db       473 AlaAsnGluSerGlySerArgGlyAspAlaSerTyrThrSerAspGluSerSerAspAsp 492

Qy       781 CTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAAT 840
                   :::|||      |||      |||   :::|||                  :::
Db       493 AspAsnAspSerAsp---SerHisAlaGlyGluAspAsp------------SerSerAsp 507

Qy       841 GACAACAAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAA 900
             |||:::         |||         |||   ||||||            :::   |||
Db       508 AspSerSerGlyAspGlyAspSerAspSerAsnGlyAspGly---------AspSerGlu 524

Qy       901 AGTGAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAA 960
             ||||||   :::         |||      :::      ||||||      |||||||||
Db       525 SerGluAspLysAspGluSerAspSerSerAspHisAspAsnSer---SerAspSerGlu 543

Qy       961 GCAGATGAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCA 1020
             :::      ::::::|||         |||:::         |||   |||      :::
Db       544 SerLysSerAspSerSerAspSerSerAspAspSerSerAspSerSerAspSerSerAsp 563

Qy      1021 AATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTG 1080
             :::      ||||||      :::      :::|||
Db       564 SerSerAspSerSerAspSerSerAspSerSerAspSer--------------------- 576

Qy      1081 ACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACC 1140
             :::      |||:::                                ::::::|||:::   :::
Db       577 SerAspSerSerAsp--------------------SerSerAspSerAsnSerSerSer 589

Qy      1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
                      ::::::      |||   :::||||||      :::         |||   |||
Db       590 AspSerSerAspSerSerGlySerSerAspSerSerAspSerSerAspThrCysAspSer 609

Qy      1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
             |||                  ||||||      :::                  |||
Db       610 SerAspSerSerAspSerSerAspSerSerAspSerSerAspSerSerAspSerSerAsp 629

Qy      1261 TCAGAGGAC 1269
             |||   |||
Db       630 SerSerAsp 632
```

```
RESULT 14
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   PIR; T34036; T34036.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;
KW   DNA-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
```

```
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;

Alignment Scores:
Pred. No.:              0.00505      Length:        1359
Score:                  157.50       Matches:       160
Percent Similarity:     33.77%       Conservative:  126
Best Local Similarity:  18.89%       Mismatches:    313
Query Match:            2.33%        Indels:        248
DB:                     1            Gaps:          36

US-09-302-812-5 (1-3814) x ATRX_CAEEL (1-1359)

Qy     283 AAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATTGACTCC 342
           |||      |||     |||:::|||||||||:::   :::   :::           :::
Db      81 LysSerArgLysArgAlaLysSerGluSerGluSerAspGluSerAspGluGluGluAsp 100

Qy     343 ATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTGGAAAAT 402
                    |||      |||                      |||    ||||||  :::
Db     101 ArgLysLysSerLysSerLysLysLysValAspGlnLysLysLysGluLysSerLys--- 119

Qy     403 GTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAACCAA 462
                 :::                  |||    ::::::   |||              :::
Db     120 ------LysLysArgThrThrSerSerSerGluAspGluAspSerAspGluGluArgGlu 137

Qy     463 CAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTG 522
           |||::::::                 |||            :::          :::|||
Db     138 GlnLysSerLysLysLysSerLysLysThrLysLysGlnThrSerSerGluSerSerGlu 157

Qy     523 GCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGTAATGCTAACATTGGT 582
              ||||||                          |||:::            ::::::
Db     158 GluSerGluGluGluArgLysValLysLysSerLysLysAsnLysGluLysSerValLys 177

Qy     583 CAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAAGAAGACAGAGACAATCAG 642
           :::         |||   :::    ||||||    |||   :::      :::   ::::::
Db     178 LysArgAlaGluThrSerGluGluSerAspGluAspGluLysProSerLysLysSerLys 197

Qy     643 CAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTAGGAGATGGGCAGGCC 702
           :::   |||            |||   |||   ::::::          :::   ::::::
Db     198 LysGlyLeuLys------LysLysAlaLysSerGlu---------SerGluSerGluSer 212

Qy     703 AGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAA 762
                    :::            |||::::::      |||                :::
Db     213 GluAspGluLysGluValLysLysSerLysLysLysSerLysLysValValLysLysGlu 232

Qy     763 CAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATT 822
              |||      |||      ||||||
Db     233 SerGluSerGluAspGluAlaProGluLysLys--------------------------- 243

Qy     823 GGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGCAGC---------CTA 873
              |||      :::      :::|||   :::   :::|||||||||
Db     244 LysThrGluLysArgLysArgSerLysThrSerSerGluGluSerSerGluSerGluLys 263

Qy     874 GGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCT------------GAGTCACCAATG 921
              |||               ||||||||||||   |||         :::    |||:::
Db     264 SerAspGluGluGluGluGluLysGluSerSerProLysProLysLysLysLysProLeu 283

Qy     922 GATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCA 981
              |||            :::      :::      |||:::|||   |||         |||
Db     284 AlaValLysLysLeuSerSerAspGluGluSerGluGluSerAspValGluValLeuPro 303

Qy     982 ------------------------------GTCTTTGATGAGCAAGATGATCGTTCCTCC 1011
                                         :::    |||    :::|||::::::
Db     304 GlnLysLysLysArgGlyAlaValThrLeuIleSerAspSerGluAspGluLysAspGln 323

Qy    1012 CAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAA 1071
           ::::::   ::::::   |||         :::   :::   :::            :::|||
Db     324 LysSerGluSerGluAlaSerAspValGluGluLysValSerLysLysLysAlaLysLys 343

Qy    1072 CGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAAGGAGAAAATAAT 1131
           :::                                                |||::::::
Db     344 Gln---------------------------------------------GluSerSer 347

Qy    1132 GCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAG 1191
                    ||||||                         :::    |||:::   |||
Db     348 GluSerGlySerAsp------------------SerSerGluGlySerIleThrValAsn 361

Qy    1192 TGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGA 1251
              |||    ||||||:::         |||   :::      |||            |||
Db     362 ---ArgLysSerLysLysLysGluLysProGluLysLysLysLysGlyIleIleMetAsp 380

Qy    1252 ATTTCCAAGTCAGAG------------GACAGAAGAAAAGAACAATGTGAAGTCAGACAT 1299
              ||||||   :::         |||   :::   :::   |||   |||
Db     381 SerSerLysLeuGlnLysGluThrIleAspAlaGluArgAlaGluLysGlu---ArgArg 399

Qy    1300 CAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAG 1359
           :::|||   |||:::|||                                 :::|||:::
Db     400 LysArgLeuGluLysLys---------------------------------GlnLysGlu 408

Qy    1360 TGGCTGGGAACTCCTATTGAGGAA------------------------------------ 1383
           :::   |||      :::||||||
Db     409 PheAsnGlyIleValLeuGluGluGlyGluAspLeuThrGluMetLeuThrGlyThrSer 428

Qy    1384 ---ATGAGAAAAATG---------------------CCTCGGTGTGGGATCCATTTGCCT 1419
              ||||||:::                                      :::
Db     429 SerGlnArgLysLeuLysSerValValLeuAspProAspSerSerThrValAspGluGlu 448

Qy    1420 TCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGA 1479
           |||   :::|||         |||         ||||||
Db     449 SerLysLysProValGluValHisAsnSerLeuValArg--------------------- 461

Qy    1480 GAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAA 1539
              :::   ||||||         |||
Db     462 ---IleLeuLysProHisGlnAlaHis--------------------------------- 469

Qy    1540 ATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA------ 1593
                    |||   :::|||               |||:::   :::|||
Db     470 ---------GlyIleGlnPheMetTyrAspCysAlaCysGluSerLeuAspArgLeuAsp 486

Qy    1594 ACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCC 1653
           |||   ||||||
Db     487 ThrGluGlySer------------------------------------------------ 490

Qy    1654 CAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTT 1713
                       ||||||   :::   :::               |||
Db     491 ------------GlyGlyIleLeuAlaHisCysMetGlyLeuGlyLysThr--------- 503

Qy    1714 ACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTATAT 1773
              ::::::   |||         ||||||                          ::::::
Db     504 LeuGlnValIleThrPheLeuHisThrValLeu--------------------MetHis 516

Qy    1774 CAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCA---AATATTTGCACC 1830
           :::   |||               ::::::   :::   :::|||   |||:::
Db     517 GluLysIleGlyGluLysCysLysArgValLeuValValValProLysAsnValIle--- 535

Qy    1831 CAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAG 1890
                 |||      |||:::         |||         :::   |||:::
Db     536 ------IleAsnTrpPheLysGluPheGlnLysTrpLeuValAsp---AsnAspGluGlu 552

Qy    1891 ATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATG 1950
           :::   ::::::   |||
Db     553 LeuAspThrIleAspValAsn--------------------------------------- 559

Qy    1951 AAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGG------TTGTTTGAAGGA 2004
                 |||      ||||||   |||         |||      |||
Db     560 ------GluLeuAspSerTyrLysThrIleGluAspArgArgArgAlaLeuLysAlaTrp 577

Qy    2005 CGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAGAGTCACA 2064
              ||||||:::   |||         :::         :::      ||||||   :::|||
Db     578 HisSerSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgIleLeuThr 597

Qy    2065 -----------------GAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGT 2106
```

```
                       :::|||||||||                       ::::::
Db        598 ValGluAspAspProLysLysLysLysProLysAsnArgAsnArgArgLeuGluLysAla 617

Qy       2107 CTTGAAGATTTT-----------------------CCAGAATGGGAAAGGTGTGAA--- 2139
                |||||||||                           |||:::      |||:::
Db        618 LysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValCysAspGlu 637

Qy       2140 ---------------------------------------------AAGCCTCTGACA 2151
                                                           |||  ||||||
Db        638 AlaHisLysLeuLysAsnAspAspSerAlaLeuSerLysCysMetValLysIleLeuThr 657

Qy       2152 CGCTTACACGTCACTTACGAGGGTACC---ATAGAAGGCAAC----------------- 2190
              :::      :::          |||||    ::::::   |||
Db        658 LysArgArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluTyrHisCys 677

Qy       2191 ---------------GGCCGAGGCATGCTA------CAGGTGGATTTTGCAAATCGTTTT 2229
                                    |||:::|||      :::   :::|||||||||||||||
Db        678 MetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaAsnArgPhe 697

Qy       2230 GTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAA------------GAAATCAGA 2277
              |||          :::       |||      ::::::             |||:::
Db        698 Val------AsnIleIleAsnArgGlyArgThrLysAspAlaSerProLeuGluValSer 715

Qy       2278 TTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAAT 2337
              |||:::                ::::::             :::    :::|||
Db        716 PheMetLysArgArgCysHisValLeuTyrAspHisLeuLysLysCysValAspArgLys 735

Qy       2338 GAGTGTCTTATTATCACAGGTACT------------------------------------ 2361
              :::       ::::::|||
Db        736 AspTyrArgValLeuThrGluAlaIleProProLysGlnGluTyrValIleAsnValArg 755

Qy       2362 ---------GAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGA 2412
                          |||    :::    |||      |||    |||      |||||||||
Db        756 GlnThrGluArgGlnCysAlaLeuTyr-AsnAlaPheLeuAsnAspIleValGly----- 773

Qy       2413 AGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCC 2472
                                                  ||||||    ::::::|||    |||||
Db        774 -------------------------------AspSerGlyLeuSerLysArgLeuLeuPr 783

Qy       2473 ATTGACGCACTTCACTTCA 2491
              |       |||       |||
Db        783 oAspTyrHisMetPheSer 789

RESULT 15
FA5_BOVIN
ID   FA5_BOVIN       STANDARD;      PRT;  2211 AA.
AC   Q28107; Q28108;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coagulation factor V precursor (Activated protein C cofactor).
GN   F5.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92147638; PubMed=1737753;
RA   Guinto E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
RT   "The complete cDNA sequence of bovine coagulation factor V.";
RL   J. Biol. Chem. 267:2971-2978(1992).
RN   [2]
RP   DISULFIDE BONDS.
RX   MEDLINE=95034740; PubMed=7947716;
RA   Xue J., Kalafatis M., Silveira J.R., Kung C., Mann K.G.;
RT   "Determination of the disulfide bridges in factor Va heavy chain.";
RL   Biochemistry 33:13109-13116(1994).
CC   -!- FUNCTION: Coagulation factor V is a cofactor that participates
CC       with factor Xa to activate prothrombin to thrombin.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC       chain, noncovalently bound. The interaction between the two chains
CC       is calcium-dependent.
CC   -!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
CC       AA repeats.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC       cofactor, factor Va (formation of a heavy chain at the N-
CC       terminus and a light chain at the C-terminus).
CC   -!- PTM: Sulfation is required for efficient thrombin cleavage and
CC       activation and for full procoagulant activity (By similarity).
CC   -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC   -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC   -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M81440; AAA30512.1; -.
DR   EMBL; M81441; AAA30513.1; -.
DR   PIR; A42580; KFBO5.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR008972; Cupredoxin.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS50022; FA58C_3; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
KW   Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW   Repeat.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   2211       COAGULATION FACTOR V.
FT   CHAIN        29    741       COAGULATION FACTOR HEAVY CHAIN (BY
FT                                SIMILARITY).
FT   PROPEP      742   1564       ACTIVATION PEPTIDE (CONNECTING REGION)
FT                                (BY SIMILARITY).
FT   CHAIN      1565   2211       COAGULATION FACTOR LIGHT CHAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       30    327       F5/8 TYPE A 1.
FT   DOMAIN       30    193       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      203    327       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      348    686       F5/8 TYPE A 2.
FT   DOMAIN      348    525       PLASTOCYANIN-LIKE 3.
FT   DOMAIN      535    686       PLASTOCYANIN-LIKE 4.
FT   DOMAIN      696   1564       B.
FT   DOMAIN     1124   1151       2 X 14 AA TANDEM REPEATS.
FT   REPEAT     1124   1137       1-1.
FT   REPEAT     1138   1151       1-2.
FT   DOMAIN     1188   1453       30 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                                [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT   REPEAT     1188   1196       2-1.
FT   REPEAT     1197   1205       2-2.
FT   REPEAT     1206   1214       2-3.
FT   REPEAT     1215   1223       2-4.
FT   REPEAT     1224   1232       2-5.
FT   REPEAT     1233   1241       2-6.
FT   REPEAT     1242   1250       2-7.
FT   REPEAT     1251   1259       2-8.
FT   REPEAT     1260   1268       2-9.
FT   REPEAT     1269   1277       2-10.
FT   REPEAT     1278   1286       2-11.
FT   REPEAT     1287   1295       2-12.
FT   REPEAT     1296   1304       2-13.
FT   REPEAT     1305   1313       2-14.
FT   REPEAT     1314   1322       2-15.
```

```
FT   REPEAT     1323   1331       2-16.
FT   REPEAT     1332   1340       2-17.
FT   REPEAT     1341   1349       2-18.
FT   REPEAT     1350   1358       2-19.
FT   REPEAT     1359   1367       2-20.
FT   REPEAT     1368   1376       2-21.
FT   REPEAT     1377   1385       2-22.
FT   REPEAT     1386   1394       2-23.
FT   REPEAT     1395   1403       2-24.
FT   REPEAT     1404   1412       2-25.
FT   REPEAT     1413   1421       2-26.
FT   REPEAT     1422   1430       2-27.
FT   REPEAT     1431   1439       2-28.
FT   REPEAT     1440   1444       2-29 (PARTIAL).
FT   REPEAT     1445   1453       2-30.
FT   DOMAIN     1569   1890       F5/8 TYPE A 3.
FT   DOMAIN     1569   1738       PLASTOCYANIN-LIKE 5.
FT   DOMAIN     1748   1890       PLASTOCYANIN-LIKE 6.
FT   DOMAIN     1894   2048       F5/8 TYPE C 1.
FT   DOMAIN     2053   2208       F5/8 TYPE C 2.
FT   SITE        741    742       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1034   1035       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1564   1565       CLEAVAGE (BY THROMBIN).
FT   DISULFID    167    193
FT   DISULFID    248    329
FT   DISULFID    499    525
FT   DISULFID    607    688
FT   DISULFID   1712   1738       PROBABLE.
FT   DISULFID   1894   2048       BY SIMILARITY.
FT   DISULFID   2053   2208       BY SIMILARITY.
FT   MOD_RES     697    697       SULFATION (POTENTIAL).
FT   MOD_RES     701    701       SULFATION (POTENTIAL).
FT   MOD_RES     730    730       SULFATION (POTENTIAL).
FT   MOD_RES    1513   1513       SULFATION (POTENTIAL).
FT   MOD_RES    1529   1529       SULFATION (POTENTIAL).
FT   MOD_RES    1537   1537       SULFATION (POTENTIAL).
FT   MOD_RES    1541   1541       SULFATION (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    587    587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    745    745       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    756    756       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    902    902       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    952    952       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1044   1044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1053   1053       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1062   1062       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1078   1078       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1094   1094       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1451   1451       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1490   1490       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1550   1550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1690   1690       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1839   1839       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1997   1997       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2196   2196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     587    592       NFTLPA -> T (IN VARIANT 2).
SQ   SEQUENCE   2211 AA;  248981 MW;  CBBF90B738667C45 CRC64;

Alignment Scores:
Pred. No.:               0.00589       Length:         2211
Score:                   157.50        Matches:        135
Percent Similarity:      32.87%        Conservative:   102
Best Local Similarity:   18.72%        Mismatches:     295
Query Match:             2.33%         Indels:         189
```

```
DB:                     1              Gaps:          28

US-09-302-812-5 (1-3814) x FA5_BOVIN (1-2211)

Qy      220 CACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATTACTACTTGGATGGAT 279
            |||             |||   |||        :::   :::::::::|||     ||||||
Db      641 HisGluAspThrLeuThrLeuPheProMetGlnGlyGluSerValThrValThrMetAsp 660

Qy      280 ACTAAAGGACCC------------------------------------------------ 291
                  |||
Db      661 AsnValGlyThrTrpMetLeuThrThrMetAsnSerAsnProArgSerLysLysLeuArg 680

Qy      291 ------------------------------------------------------------ 291

Db      681 LeuArgPheArgAspAlaLysCysIleArgAsnAspAspAspAspSerTyrGluIleIle 700

Qy      292 ---------------AAGACAGCTGAATCAGAAAGTAAAGAAAACAACAATACAAGAATT 336
                           ||||||   :::        |||   :::::::::::    |||
Db      701 TyrGluProSerGlySerThrAlaMetThrThrLysLysIleHisAspSerSerGluIle 720

Qy      337 GACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGGAAAAATTG 396
            :::         :::           |||:::   |||      ::::::      :::
Db      721 GluAspGluAsnAspAla---------AspSerAspTyrGlnAspGluLeuAlaLeuIle 737

Qy      397 ------------------------------------GAAAATGTTCCTCAGCTAAATCTT 420
                                                :::         :::|||||||||
Db      738 LeuGlyLeuArgSerPheArgAsnSerSerLeuAsnGlnGluLysAspGluLeuAsnLeu 757

Qy      421 GATAAATCACCCACAGAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTG 480
                              ||||||   |||:::::::::                |||:::
Db      758 ThrAlaLeuAlaLeuGluLysAspSerGluPheIleProProSerAlaAsnArgSerLeu 777

Qy      481 TGCAAGTGGCAGAATGAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAG--------- 531
                         :::         |||      :::|||:::|||
Db      778 AspSerAsnSerSerSerArgSerHisValSerArgLeuIleAlaLysAsnPheAlaGlu 797

Qy      532 -----------------------------CCTCCCGCGGGGACTCCGCTACCAAAGCAG 561
                                         |||   ||||||:::||||||
Db      798 SerLeuLysThrLeuLeuHisLeuGluAlaProAlaAlaGlySerProLeuGluHisAla 817

Qy      562 CTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCAT 621
                  :::   |||         :::|||        :::||||||
Db      818 GlyLeuAspLysAsnSerAlaLeuAsnProProMetAlaGluHisSerSerProTyrSer 837

Qy      622 GAAGAAGACAGAGACAATCAGCAGTTT------CTTACACCTATAAAACTT--------- 666
            |||:::   |||::::::               :::|||   :::   |||
Db      838 GluAspProArgGluAspHisProLeuSerAspValThrGlyValSerLeuLeuProPhe 857

Qy      667 -----------GCAAATACAAAGCCAACA-----------------GTAGGAGATGGG 696
                       |||   ||||||                       ||||||   |||
Db      858 GlyThrGlyPheLysAsnArgLysProAlaLysHisGlnArgPheGlnValGlyArgGly 877

Qy      697 CAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGC---CAGTCTGTGAAAGACTGTACA 753
            ||||||      :::   |||   |||   :::|||      :::   |||      |||
Db      878 GlnAlaAlaLysHis---LysPheSerGlnThrArgPheProAlaHisLysThrArgThr 896

Qy      754 GGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTTGGTGCC 813
                  |||:::                             |||   |||:::   :::
Db      897 ArgLeuSerGlnAspAsnSerSerSerArgMetGlyProTrpGluAspIleProSer 916

Qy      814 GAGGACATTGGAACTGGACCAAAAAATGACAACAAA---TTGACTGGACAAGAAAGCAGC 870
            :::   :::            |||:::      |||   |||   |||:::
Db      917 AspLeuLeuLeuLeuGlnGlnLysAspProTyrLysIleLeuAsnGlyGluTrpHisLeu 936

Qy      871 CTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTAGAC 930
            :::   :::         :::|||         :::                  |||:::
Db      937 ValSerGluLysGlySerTyrGluIleIleGlnAspAlaAsnGluAsnLysThrValAsn 956

Qy      931 AACTCGAGAAACAGTTGTCAA--------------------------------------- 951
                  ||||||   |||
Db      957 LysLeuProAsnSerProGlnAsnAspSerArgThrTrpGlyGluAsnIleProPheLys 976

Qy      952 GATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTT------------------------ 987
```

```
                 :::|||          :::                   |||    |||
Db        977 AsnSerHisGlyLysGlnSerGlyHisProThrPheLeuValThrArgArgLysProLeu 996

Qy        988 ---GATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCA 1044
                 |||    ||||||    |||:::|||:::           |||          :::
Db        997 GlnAspArgGlnAspArgArgAsnSerArgLeuLysGluGlyLeuProLeuIleArgThr 1016

Qy       1045 AGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTG 1104
              |||              :::   :::        |||               |||
Db       1017 ArgArgLysLysLysGluGluLysProAlaTyrHisValProLeuSerProArgSerPhe 1036

Qy       1105 CATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCT 1164
              |||              ||||||   ||||||      ||||||               |||
Db       1037 His---ProLeuArgGlyGluValAsnAlaSerPheSerAspArgArgHisAsnHisSer 1055

Qy       1165 ---------------GGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCC------ 1203
                                 |||    :::||||||:::::::::         :::
Db       1056 LeuLeuLeuHisAlaSerAsnGluThrSerLeuSerIleAspLeuAsnGlnThrPhePro 1075

Qy       1204 AAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCA 1263
                    :::          :::    :::   ||||||
Db       1076 SerMetAsnLeuSerLeuAlaAlaSerLeuProAspHis--------------------- 1088

Qy       1264 GAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCA 1323
                             :::|||                       ::::::          |||
Db       1089 ---------------AspGlnThrSerProAsnAspThrThrSerGlnThrSerSerPro 1103

Qy       1324 AAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAA 1383
                    :::   |||   :::   ||||||:::   :::          ||||||::::::
Db       1104 ProAspLeuTyrProThrValSerProGluGluHisTyrGlnIlePheProIleGlnAsp 1123

Qy       1384 ATGAGAAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGA---CCATCTGCAAGT 1440
                          |||            ||||||   |||   |||          :::
Db       1124 ------SerAspProThrHisSerThrThrAlaProSerAsnArgSerProAspProThr 1141

Qy       1441 CACACAGTGACTGTT---------CGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAA 1491
              |||:::   |||            |||            :::      ::::::|||
Db       1142 HisSerThrThrAlaProSerAsnArgSerProProThrGlnProSerGlnIleProAsn 1161

Qy       1492 CCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCG 1551
                                         :::|||   ::::::      :::|||
Db       1162 ---------------------------TyrAspLeuArgAsnArgAlaIleProThrAsp 1172

Qy       1552 GAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGG 1611
                       ::::::|||
Db       1173 ValSerGlnIlePhePro-----------------------------SerLeuGluLeu 1182

Qy       1612 GAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCG 1671
              |||:::   |||||||||                ::::::|||
Db       1183 GluValTrpGlnThrAlaThrSerLeuAspLeuSerGlnPro------------------ 1196

Qy       1672 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGGTTGATTTC 1731
                                                       :::      |||
Db       1197 ------------------------------------------SerIleSerProAspLeu 1202

Qy       1732 TGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGAC 1791
                       |||             |||          :::|||:::   ||||||
Db       1203 GlyGlnMetAlaLeuSerProAspProGlyGln---------GluSerLeuSerProAsp 1219

Qy       1792 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA---CCACTCCTG 1848
              :::   :::   :::|||         |||::::::      :::   :::   |||   |||
Db       1220 LeuGlyGlnThrSerLeuSer---ProAspLeuSerGlnGluSerLeuSerProAspLeu 1238

Qy       1849 AAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCT 1908
                 |||         |||         |||||||||      ::::::      |||
Db       1239 GlyGlnThrAla---LeuSerProAspProSerGlnGluSerLeuSerProAspLeuGly 1257

Qy       1909 AATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGT 1968
                                 |||                   |||:::   |||
Db       1258 GlnThrAlaLeuSerProAspPro---------------------SerGlnGluSerLeu 1270

Qy       1969 TACCCAGACATTAACTTCAATCGGTTGTTT------GAAGGACGTTCATCAAGGAAACCA 2022
                 ||||||:::               |||         |||:::   |||      |||
Db       1271 SerProAspLeuGlyGlnThrAlaLeuSerProAspProGlyGlnGluSerLeuSerPro 1290

Qy       2023 GAA 2025
              :::
Db       1291 Asp 1291
```

Search completed: May 26, 2004, 16:55:49
Job time : 221.908 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 26, 2004, 16:13:47 ; Search time 178.048 Seconds
                                          (without alignments)
                                          13517.524 Million cell updates/sec

Title:          US-09-302-812-5
Perfect score:  6773
Sequence:       1 gggggactgtgtgctgcggg..........aatcatttgtcagaaaaaaa 3814

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150053_5676/app_query.fasta_1.12437
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_640_@runat_26052004_150053_5676 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5164	76.2	968	11	O88622	O88622 mus musculu
2	5098.5	75.3	961	11	Q80YQ6	Q80yq6 mus musculu
3	4869.5	71.9	920	11	Q8CB72	Q8cb72 mus musculu
4	4749	70.1	972	11	Q9QYM2	Q9qym2 rattus norv
5	4456	65.8	976	4	Q86W56	Q86w56 homo sapien
6	4451	65.7	976	4	Q7Z742	Q7z742 homo sapien
7	4395	64.9	976	4	Q9Y4W7	Q9y4w7 homo sapien
8	4332.5	64.0	977	6	O02776	O02776 bos taurus
9	1074.5	15.9	768	5	O46043	O46043 drosophila
10	1069.5	15.8	723	5	Q960N8	Q960n8 drosophila
11	568	8.4	548	10	Q9SKB3	Q9skb3 arabidopsis
12	528	7.8	522	10	Q8VYA1	Q8vya1 arabidopsis
13	458	6.8	485	5	Q9N5L4	Q9n5l4 caenorhabdi
14	451	6.7	781	5	Q867X0	Q867x0 caenorhabdi
15	450.5	6.7	764	5	Q19637	Q19637 caenorhabdi
16	328.5	4.9	368	5	Q86GI4	Q86gi4 toxoplasma
17	328	4.8	364	10	Q9SKB4	Q9skb4 arabidopsis
18	262.5	3.9	747	4	Q86V50	Q86v50 homo sapien
19	190	2.8	931	5	Q9VF92	Q9vf92 drosophila
20	188.5	2.8	930	5	Q8IHH1	Q8ihh1 drosophila
21	183	2.7	1194	11	Q8C0H4	Q8c0h4 mus musculu
22	183	2.7	1194	11	Q8BMI0	Q8bmi0 mus musculu
23	182	2.7	571	5	Q8MTN7	Q8mtn7 trichinella
24	179.5	2.7	1193	11	Q8BTQ4	Q8btq4 mus musculu
25	177.5	2.6	3166	5	Q9W3Z0	Q9w3z0 drosophila
26	177	2.6	5322	5	Q9VPL9	Q9vpl9 drosophila
27	176.5	2.6	888	12	Q8JKL5	Q8jkl5 heliothis z
28	176	2.6	8081	5	Q7Z120	Q7z120 caenorhabdi
29	173	2.6	1537	4	Q96NW7	Q96nw7 homo sapien
30	172	2.5	508	12	Q9E234	Q9e234 helicoverpa
31	169	2.5	700	4	Q9BUV3	Q9buv3 homo sapien
32	167	2.5	510	12	Q99GU9	Q99gu9 helicoverpa
33	167	2.5	1169	16	Q7ZAJ1	Q7zaj1 staphylococ
34	167	2.5	1758	4	Q7Z6E8	Q7z6e8 homo sapien
35	166	2.5	1616	4	Q96PH3	Q96ph3 homo sapien
36	166	2.5	1792	4	Q7Z6E9	Q7z6e9 homo sapien
37	165	2.4	498	5	Q8MTN8	Q8mtn8 trichinella
38	165	2.4	1192	5	O96127	O96127 plasmodium
39	164	2.4	766	11	Q8C1D8	Q8c1d8 mus musculu
40	163	2.4	532	10	Q9T069	Q9t069 arabidopsis
41	163	2.4	691	4	Q96QF7	Q96qf7 homo sapien
42	163	2.4	2392	5	Q8I0K1	Q8i0k1 caenorhabdi
43	163	2.4	2407	5	Q7YSH5	Q7ysh5 caenorhabdi
44	163	2.4	2677	4	Q7Z333	Q7z333 homo sapien
45	162.5	2.4	1508	5	Q86J39	Q86j39 dictyosteli

ALIGNMENTS

```
RESULT 1
O88622
ID   O88622      PRELIMINARY;      PRT;   968 AA.
AC   O88622;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   PARG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380098; PubMed=10449915;
RA   Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;
RT   "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT   human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT   hybridization.";
RL   Cytogenet. Cell Genet. 85:269-270(1999).
DR   EMBL; AF079557; AAC28735.1; -.
DR   MGD; MGI:1347094; Parg.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
```

```
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   968 AA;  109169 MW;  3981DA4809E1E007 CRC64;

Alignment Scores:
Pred. No.:              0             Length:        968
Score:                  5164.00       Matches:       968
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            76.24%        Indels:        0
DB:                     11            Gaps:          0

US-09-302-812-5 (1-3814) x O88622 (1-968)

Qy         28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAAGCGCGCTGGGGCGCCGCTGGAACT 87
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20

Qy         88 TCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTCGAC 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeuAsp 40

Qy        148 CCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCGGGG 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60

Qy        208 CGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATTACT 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80

Qy        268 ACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAACAAT 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsnAsn 100

Qy        328 ACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTG 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120

Qy        388 GAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGTTCA 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140

Qy        448 CAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAGCAT 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160

Qy        508 GCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTTAGT 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180

Qy        568 AATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAAGAA 627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200

Qy        628 GACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACAGTA 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220

Qy        688 GGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGAC 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240

Qy        748 TGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGATGTT 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAspVal 260

Qy        808 GGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAAAGC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280

Qy        868 AGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300

Qy        928 GACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTCTTT 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320

Qy        988 GATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGA 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340

Qy       1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360

Qy       1108 TTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380

Qy       1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGATTCT 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400

Qy       1228 AAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGT 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420

Qy       1288 GAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCT 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440

Qy       1348 CCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460

Qy       1408 ATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTT 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480

Qy       1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAAC 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500

Qy       1528 AAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGT 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520

Qy       1588 GAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACA 1647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540

Qy       1648 CGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGG 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560

Qy       1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT 1767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580

Qy       1768 TTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGC 1827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600

Qy       1828 ACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAA 1887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620

Qy       1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640

Qy       1948 ATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGT 2007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660
```

```
Qy       2008 TCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAG 2067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680

Qy       2068 AAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGG 2127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700

Qy       2128 GAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAAGGC 2187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720

Qy       2188 AACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACT 2247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyGlyValThr 740

Qy       2248 GGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCA 2307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760

Qy       2308 CGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAG 2367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780

Qy       2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATGGG 2427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800

Qy       2428 AGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820

Qy       2488 TTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCT 2547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840

Qy       2548 TACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGA 2607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860

Qy       2608 AACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAGATCCTG 2667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880

Qy       2668 GCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAGTTGATG 2727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900

Qy       2728 AGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTG 2787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920

Qy       2788 TACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA 2847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940

Qy       2848 GACATCAAGCTTTATCCATTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGAC 2907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960

Qy       2908 ATGCCAGGACAGAAGGCAGGCACC 2931
              ||||||||||||||||||||||||
Db        961 MetProGlyGlnLysAlaGlyThr 968
```

```
RESULT 2
Q80YQ6
ID   Q80YQ6      PRELIMINARY;      PRT;   961 AA.
AC   Q80YQ6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Poly (ADP-ribose) glycohydrolase.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050892; AAH50892.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   961 AA;  108580 MW;  1558F120C367CF63 CRC64;

Alignment Scores:
Pred. No.:               0             Length:         961
Score:                   5098.50       Matches:        957
Percent Similarity:      99.69%        Conservative:   1
Best Local Similarity:   99.58%        Mismatches:     2
Query Match:             75.28%        Indels:         1
DB:                      11            Gaps:           1

US-09-302-812-5 (1-3814) x Q80YQ6 (1-961)

Qy         28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
              ||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy         85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60

Qy        205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

Qy        265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAAC 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100

Qy        325 AATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAG 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLys 120

Qy        385 GTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGT 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140

Qy        445 TCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAG 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLys 160

Qy        505 CATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTT 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 HisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeu 180

Qy        565 AGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAA 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGlu 200

Qy        625 GAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACA 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThr 220

Qy        685 GTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ValGlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLys 240
```

```
Qy       745 GACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGAT 804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AspCysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAsp 260

Qy       805 GTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAA 864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGlu 280

Qy       865 AGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGAT 924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 SerSerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAsp 300

Qy       925 GTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC 984
             |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProVal 320

Qy       985 TTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCA 1044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 PheAspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAla 340

Qy      1045 AGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTG 1104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 ArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeu 360

Qy      1105 CATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCT 1164
             |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       361 HisPheGlnPheGluGlyGluAsnAsnValGlyThrSerAspLeuAsnAlaLysProSer 380

Qy      1165 GGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGAT 1224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 GlyAsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAsp 400

Qy      1225 TCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAA 1284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 SerLysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGln 420

Qy      1285 TGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTC 1344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CysGluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeu 440

Qy      1345 CCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGT 1404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ProProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCys 460

Qy      1405 GGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGAC 1464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 GlyIleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAsp 480

Qy      1465 CTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGAT 1524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 LeuLeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAsp 500

Qy      1525 AACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAAT 1584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       501 AsnLysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsn 520

Qy      1585 GGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTC 1644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 GlyGluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPhe 540

Qy      1645 ACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAA 1704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 ThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLys 560

Qy      1705 TGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAA 1764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 TrpAspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGln 580

Qy      1765 CATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATT 1824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 HisLeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIle 600

Qy      1825 TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAG 1884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CysThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGln 620

Qy      1885 GAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCC 1944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       621 GluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAla 640

Qy      1945 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGA 2004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       641 LysMetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGly 660

Qy      2005 CGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACA 2064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ArgSerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThr 680

Qy      2065 GAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAA 2124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       681 GluLysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGlu 700

Qy      2125 TGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAA 2184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       701 TrpGluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGlu 720

Qy      2185 GGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTG 2244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GlyAsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyGlyVal 740

Qy      2245 ACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTT 2304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       741 ThrGlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleVal 760

Qy      2305 TCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAA 2364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 SerArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGlu 780

Qy      2365 CAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGAT 2424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAsp 800

Qy      2425 GGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTT 2484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       801 GlySerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeu 820

Qy      2485 CACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAG 2544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 HisPheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLys 840

Qy      2545 GCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACG 2604
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AlaTyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThr 860

Qy      2605 GGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAGATC 2664
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       861 GlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIle 880

Qy      2665 CTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAGTTG 2724
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 LeuAlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeu 900

Qy      2725 ATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAA 2784
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 MetArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLys 920

Qy      2785 GTGTACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGA 2844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       921 ValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGly 940

Qy      2845 CCAGACATCAAGCTTTATCCATTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACT 2904
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 ProAspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThr 960

Qy      2905 GAC 2907
             |||
```

```
Db        961 Asp 961

RESULT 3
Q8CB72
ID   Q8CB72       PRELIMINARY;      PRT;   920 AA.
AC   Q8CB72;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly.
GN   PARG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Bone;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK036656; BAC29519.1; -.
DR   MGD; MGI:1347094; Parg.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   920 AA;  103799 MW;  379E779E121B09A1 CRC64;

Alignment Scores:
Pred. No.:              0            Length:         920
Score:                  4869.50      Matches:        915
Percent Similarity:     99.67%       Conservative:   2
Best Local Similarity:  99.46%       Mismatches:     2
Query Match:            71.90%       Indels:         1
DB:                     11           Gaps:           1

US-09-302-812-5 (1-3814) x Q8CB72 (1-920)

Qy         28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
              |||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy         85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60

Qy        205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

Qy        265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAAC 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100

Qy        325 AATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAG 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLys 120

Qy        385 GTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGT 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140

Qy        445 TCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAG 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLys 160

Qy        505 CATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTT 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 HisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeu 180

Qy        565 AGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAA 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGlu 200

Qy        625 GAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACA 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThr 220

Qy        685 GTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ValGlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLys 240

Qy        745 GACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGAT 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspCysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAsp 260

Qy        805 GTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAA 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGlu 280

Qy        865 AGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGAT 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 SerSerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAsp 300

Qy        925 GTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC 984
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProVal 320

Qy        985 TTTGATGAGCAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCA 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 PheAspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAla 340

Qy       1045 AGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTG 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeu 360

Qy       1105 CATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCT 1164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 HisPheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSer 380

Qy       1165 GGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGGGAT 1224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GlyAsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAsp 400

Qy       1225 TCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAA 1284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerLysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGln 420

Qy       1285 TGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTC 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CysGluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeu 440

Qy       1345 CCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGT 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCys 460

Qy       1405 GGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGAC 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GlyIleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAsp 480

Qy       1465 CTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGAT 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LeuLeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAsp 500

Qy       1525 AACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAAT 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AsnLysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsn 520

Qy       1585 GGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTC 1644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    521 GlyGluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPhe 540

Qy   1645 ACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAA 1704
Db    541 ThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLys 560

Qy   1705 TGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAA 1764
Db    561 TrpAspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGln 580

Qy   1765 CATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATT 1824
Db    581 HisLeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIle 600

Qy   1825 TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAG 1884
Db    601 CysThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGln 620

Qy   1885 GAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCC 1944
Db    621 GluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAla 640

Qy   1945 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGA 2004
Db    641 LysMetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGly 660

Qy   2005 CGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACA 2064
Db    661 ArgSerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThr 680

Qy   2065 GAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAA 2124
Db    681 GluLysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGlu 700

Qy   2125 TGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAA 2184
Db    701 TrpGluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGlu 720

Qy   2185 GGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTG 2244
Db    721 GlyAsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyGlyVal 740

Qy   2245 ACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTT 2304
Db    741 ThrGlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleVal 760

Qy   2305 TCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAA 2364
Db    761 SerArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGlu 780

Qy   2365 CAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGAT 2424
Db    781 GlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAsp 800

Qy   2425 GGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTT 2484
Db    801 GlySerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeu 820

Qy   2485 CACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAG 2544
Db    821 HisPheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLys 840

Qy   2545 GCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACG 2604
Db    841 AlaTyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThr 860

Qy   2605 GGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAGATC 2664
Db    861 GlyAsnTrpGlyCysGlyAlaPheGlyGlyAspValArgLeuLysAlaLeuIleGlnIle 880

Qy   2665 CTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAGTTG 2724
Db    881 LeuAlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeu 900

Qy   2725 ATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAA 2784
Db    901 MetArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyGlu 920
```

RESULT 4
Q9QYM2

```
ID   Q9QYM2     PRELIMINARY;     PRT;   972 AA.
AC   Q9QYM2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   PARG.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BUF; TISSUE=Colon;
RA   Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA   Aoki Y., Nakgama H., Sugimura T.;
RT   "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RT   (Parg).";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB019366; BAA87901.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   972 AA;  109002 MW;  3DED7885684F59E9 CRC64;
```

Alignment Scores:

Pred. No.:	0	Length:	972
Score:	4749.00	Matches:	894
Percent Similarity:	94.65%	Conservative:	26
Best Local Similarity:	91.98%	Mismatches:	48
Query Match:	70.12%	Indels:	4
DB:	11	Gaps:	2

US-09-302-812-5 (1-3814) x Q9QYM2 (1-972)

```
Qy     28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
Db      1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy     85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
Db     21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnLysArgValLeu 40

Qy    145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
Db     41 AspProLysAspAlaProValGlnPheArgValProProSerSerSerAlaCysValSer 60

Qy    205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTCAAACAAAAGACTATT 264
Db     61 GlyArgAlaGlyProHisArgGlySerValThrSerPheValPheLysGlnLysProIle 80

Qy    265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAAAGTAAAGAAAACAAC 324
Db     81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100

Qy    325 AATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAG 384
Db    101 AsnThrArgThrAspProMetMetSerSerValGlnLysAspAsnPheTyrProHisLys 120

Qy    385 GTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACAGAAAAGAGT 444
Db    121 ValGluLysLeuGlyAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140

Qy    445 TCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAATGAAGGGAAG 504
```

```
Db    141 ThrProTyrLeuAsnGlnGlnGlnThrAlaGlyValCysLysTrpHisSerAlaGlyGlu 160

Qy    505 CATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCAAAGCAGCTT 564
Db    161 ArgAlaGluGlnLeuSerAlaSerGluProSerAlaValThrGlnAlaProLysGlnLeu 180

Qy    565 AGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACAGATCATGAA 624
Db    181 SerAsnAlaAsnIleAspGlnSerProProThrAspGlyHisSerAspThrAspHisGlu 200

Qy    625 GAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCAACA 684
Db    201 GluAspArgAspAsnGlnGlnPheLeuThrProValLysLeuAlaAsnAlaLysGlnThr 220

Qy    685 GTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAA 744
Db    221 ValGlyAspGlyGlnAlaArgSerAsnCysLysCysSerAlaSerCysGlnCysGlyGln 240

Qy    745 GACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCAGAT 804
Db    241 AspCysAlaGlyCysGlnArgGluGluAlaAspValIleProGluSerProLeuSerAsp 260

Qy    805 GTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAACAAATTGACTGGACAAGAA 864
Db    261 ValGlyAlaGluAspIleGlyThrGlySerLysAsnAspAsnLysLeuThrGlyGlnGlu 280

Qy    865 AGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGAT 924
Db    281 SerGlyLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAsp 300

Qy    925 GTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC 984
Db    301 ValAspAsnSerLysThrSerCysGlnAspSerGluAlaAspGluGluAlaSerProVal 320

Qy    985 TTTGATGAG---------CAAGATGATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGT 1035
Db    321 PheAspGluGlnAspAspGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSer 340

Qy   1036 TGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAA 1095
Db    341 ArgGlnAlaArgGluValAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGlu 360

Qy   1096 GTTAGATTGCATTTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCC 1155
Db    361 IleArgLeuHisPheGlnPheGluGlyGlySerAsnAlaGlyThrSerAspLeuAsnAla 380

Qy   1156 AAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGA 1215
Db    381 LysProSerGlyAsnSerSerSerLeuAsnValAspGlyArgSerSerLysGlnHisGly 400

Qy   1216 AAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGACAGAAGA 1275
Db    401 LysArgAspSerLysIleThrAspHisPheValArgIleProLysSerGluAspLysArg 420

Qy   1276 AAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATACATCCCA 1335
Db    421 LysGluGlnCysGluValArgHisGlnArgAlaGluArgLysIleProLysTyrValPro 440

Qy   1336 CCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATG 1395
Db    441 ProAsnLeuProProAspLysLysTrpLeuGlyThrProIleGluGluMetArgLysMet 460

Qy   1396 CCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTGACTGTT 1455
Db    461 ProArgCysGlyValArgLeuProLeuLeuArgProSerAlaSerHisThrValThrVal 480

Qy   1456 CGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAACACATTACAAAGAT 1515
Db    481 ArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAsp 500

Qy   1516 TTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAA 1575
Db    501 LeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGlu 520

Qy   1576 GATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTC 1635
Db    521 AspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeu 540

Qy   1636 AACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATAT 1695
Db    541 AsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyr 560

Qy   1696 TCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCA 1755
Db    561 SerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAla 580

Qy   1756 GAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTG 1815
Db    581 GluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeu 600

Qy   1816 CCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACG 1875
Db    601 ProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThr 620

Qy   1876 ATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGA 1935
Db    621 MetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArg 640

Qy   1936 CGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTG 1995
Db    641 ArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeu 660

Qy   1996 TTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGA 2055
Db    661 PheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArg 680

Qy   2056 AGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTGAAGAT 2115
Db    681 ArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAsp 700

Qy   2116 TTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGT 2175
Db    701 PheProGluTrpGluArgCysAspLysProLeuThrArgLeuHisValThrTyrGluGly 720

Qy   2176 ACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGA 2235
Db    721 ThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGly 740

Qy   2236 GGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAA 2295
Db    741 GlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGlu 760

Qy   2296 TTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACA 2355
Db    761 LeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThr 780

Qy   2356 GGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGC 2415
Db    781 GlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSer 800

Qy   2416 CATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATT 2475
Db    801 HisGluAspGlySerGluLysAspAspTrpGlnArgCysCysThrGluIleValAlaIle 820

Qy   2476 GACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAG 2535
Db    821 AspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGlu 840

Qy   2536 CTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCA 2595
Db    841 LeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProProGluAsnLeuSerAla 860

Qy   2596 GTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTA 2655
Db    861 ValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeu 880
```

```
Qy       2656 ATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGAC 2715
Db        881 IleGlnLeuLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAsp 900

Qy       2716 TCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGAT 2775
Db        901 SerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAsn 920

Qy       2776 GTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCC 2835
Db        921 ValGlyLysValTyrArgLeuLeuLeuArgTyrTyrArgGluGluCysArgAspCysSer 940

Qy       2836 ACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTTGAGTCAAGTGCA 2895
Db        941 SerProGlyProAspThrLysLeuTyrProPheIleTyrHisAlaAlaGluSerSerAla 960

Qy       2896 GAGACCACTGACATGCCAGGACAGAAGGCAGGCACC 2931
Db        961 GluThrSerAspGlnProGlyGlnArgThrGlyThr 972
```

RESULT 5
Q86W56

```
ID   Q86W56      PRELIMINARY;      PRT;   976 AA.
AC   Q86W56;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to poly (ADP-ribose) glycohydrolase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050560; AAH50560.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   976 AA;  111109 MW;  D6646353C6D0180E CRC64;

Alignment Scores:
Pred. No.:               0              Length:         976
Score:                   4456.00        Matches:        846
Percent Similarity:      91.61%         Conservative:   49
Best Local Similarity:   86.59%         Mismatches:     72
Query Match:             65.79%         Indels:         10
DB:                      4              Gaps:           6

US-09-302-812-5 (1-3814) x Q86W56 (1-976)

Qy         28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAla--- 19

Qy         85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
Db         20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy        145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
Db         40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy        205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
Db         60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy        265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAA---------AGTAAA 315
Db         80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy        316 GAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375
Db        100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy        376 CCACATAAGGTGGAAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA 435
Db        120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139

Qy        436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
Db        140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy        496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
Db        160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy        556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACA 615
Db        180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy        616 GATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675
Db        200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219

Qy        676 AAGCCAACAGTAGGAGATGGGCAGGCCAGA---------AGCAACTGTAAGTGCAGTGGA 726
Db        220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy        727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786
Db        240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy        787 GAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAAC 846
Db        260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279

Qy        847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAG 906
Db        280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy        907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGAT 966
Db        300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy        967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCCTCCCAAACAGCAAAT 1023
Db        320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy       1024 AAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083
Db        340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy       1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140
Db        360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy       1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
Db        380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy       1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
Db        400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419

Qy       1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
Db        420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy       1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
Db        440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459
```

```
Qy       1381 GAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440
Db        460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy       1441 CACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
Db        480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499

Qy       1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAAC 1560
Db        500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy       1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATT 1620
Db        520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy       1621 CAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAA 1680
Db        540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy       1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740
Db        560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy       1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800
Db        580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy       1801 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 1860
Db        600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy       1861 AATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920
Db        620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy       1921 TGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980
Db        640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy       1981 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTC 2040
Db        660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy       2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100
Db        680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy       2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160
Db        700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy       2161 GTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220
Db        720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy       2221 AATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTT 2280
Db        740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy       2281 TTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340
Db        760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy       2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400
Db        780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

Qy       2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACG 2460
Db        800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgArgCysThr 819

Qy       2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520
Db        820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy       2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCT 2580
Db        840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy       2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640
Db        860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy       2641 AGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTAT 2700
Db        880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy       2701 TTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760
Db        900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy       2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820
Db        920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy       2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880
Db        940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy       2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAAGGCAGGCACC 2931
Db        960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 6
Q7Z742
ID   Q7Z742       PRELIMINARY;      PRT;    976 AA.
AC   Q7Z742;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC052966; AAH52966.1; -.
```

```
KW   Hypothetical protein.
SQ   SEQUENCE   976 AA;  111090 MW;  47C8EE1826C4A74A CRC64;

Alignment Scores:
Pred. No.:              0              Length:         976
Score:                  4451.00        Matches:        845
Percent Similarity:     91.50%         Conservative:   49
Best Local Similarity:  86.49%         Mismatches:     73
Query Match:            65.72%         Indels:         10
DB:                     4              Gaps:           6

US-09-302-812-5 (1-3814) x Q7Z742 (1-976)

Qy     28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
Db      1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAla--- 19

Qy     85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
Db     20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy    145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
Db     40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy    205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
Db     60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy    265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAA---------AGTAAA 315
Db     80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy    316 GAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375
Db    100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy    376 CCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA 435
Db    120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139

Qy    436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
Db    140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy    496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
Db    160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy    556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACA 615
Db    180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy    616 GATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675
Db    200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219

Qy    676 AAGCCAACAGTAGGAGATGGGCAGGCCAGA---------AGCAACTGTAAGTGCAGTGGA 726
Db    220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy    727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786
Db    240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy    787 GAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAAC 846
Db    260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279

Qy    847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAG 906
Db    280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy    907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGAT 966
Db    300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy    967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCCTCCCAAACAGCAAAT 1023
Db    320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy   1024 AAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083
Db    340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy   1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140
Db    360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy   1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
Db    380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy   1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
Db    400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419

Qy   1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
Db    420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy   1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
Db    440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459

Qy   1381 GAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440
Db    460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy   1441 CACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
Db    480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499

Qy   1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAAC 1560
Db    500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy   1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATT 1620
Db    520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy   1621 CAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAA 1680
Db    540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy   1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740
Db    560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy   1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800
Db    580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy   1801 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 1860
Db    600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy   1861 AATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920
Db    620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy   1921 TGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980
Db    640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy   1981 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTC 2040
```

```
Db     660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy    2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100
Db     680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy    2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160
Db     700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy    2161 GTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220
Db     720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy    2221 AATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTT 2280
Db     740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy    2281 TTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340
Db     760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy    2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400
Db     780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

Qy    2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACG 2460
Db     800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgHisCysThr 819

Qy    2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520
Db     820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy    2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCT 2580
Db     840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy    2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640
Db     860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy    2641 AGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTAT 2700
Db     880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy    2701 TTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760
Db     900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy    2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820
Db     920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy    2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880
Db     940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy    2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAAGGCAGGCACC 2931
Db     960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

RESULT 7

```
Q9Y4W7
ID   Q9Y4W7      PRELIMINARY;      PRT;    976 AA.
AC   Q9Y4W7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   HPARG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380098; PubMed=10449915;
RA   Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;
RT   "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT   human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT   hybridization.";
RL   Cytogenet. Cell Genet. 85:269-270(1999).
DR   EMBL; AF005043; AAB61614.1; -.
DR   Genew; HGNC:8605; PARG.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   976 AA;  110903 MW;  42A0F51466FCE7DC CRC64;

Alignment Scores:
Pred. No.:                0             Length:          976
Score:                    4395.00       Matches:         837
Percent Similarity:       90.99%        Conservative:    52
Best Local Similarity:    85.67%        Mismatches:      78
Query Match:              64.89%        Indels:          10
DB:                       4             Gaps:            6

US-09-302-812-5 (1-3814) x Q9Y4W7 (1-976)

Qy      28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAAGCG---CGCTGGGGCGCCGCTGGA 84
Db       1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19

Qy      85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
Db      20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy     145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
Db      40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy     205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
Db      60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy     265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAA---------AGTAAA 315
Db      80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy     316 GAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375
Db     100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy     376 CCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA 435
Db     120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139

Qy     436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
Db     140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy     496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
Db     160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy     556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACA 615
Db     180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy     616 GATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675
Db     200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
```

```
Qy    676 AAGCCAACAGTAGGAGATGGGCAGGCCAGA---------AGCAACTGTAAGTGCAGTGGA 726
Db    220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy    727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786
Db    240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy    787 GAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAAC 846
Db    260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279

Qy    847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAG 906
Db    280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy    907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGAT 966
Db    300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy    967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCCTCCCAAACAGCAAAT 1023
Db    320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy   1024 AAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083
Db    340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy   1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140
Db    360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy   1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
Db    380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy   1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
Db    400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419

Qy   1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
Db    420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy   1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
Db    440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459

Qy   1381 GAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440
Db    460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy   1441 CACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
Db    480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499

Qy   1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAAC 1560
Db    500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy   1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATT 1620
Db    520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy   1621 CAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAA 1680
Db    540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy   1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740
Db    560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy   1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800
Db    580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy   1801 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 1860
Db    600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy   1861 AATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920
Db    620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy   1921 TGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980
Db    640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy   1981 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTC 2040
Db    660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy   2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100
Db    680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy   2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160
Db    700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy   2161 GTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220
Db    720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy   2221 AATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTT 2280
Db    740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy   2281 TTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340
Db    760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy   2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400
Db    780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

Qy   2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACG 2460
Db    800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

Qy   2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520
Db    820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy   2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCT 2580
Db    840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy   2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640
Db    860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy   2641 AGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTAT 2700
Db    880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy   2701 TTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760
Db    900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy   2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820
Db    920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy   2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy      2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAAGGCAGGCACC 2931
             |||||||||  |||||||||   |||      ||||||:::   ||||||
Db       960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 8
O02776
ID   O02776      PRELIMINARY;      PRT;    977 AA.
AC   O02776;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   BPARG.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97277328; PubMed=9115250;
RA   Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT   "Isolation and characterization of the cDNA encoding bovine poly(ADP-
RT   ribose) glycohydrolase.";
RL   J. Biol. Chem. 272:11895-11901(1997).
DR   EMBL; U78975; AAB53370.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   977 AA;  110837 MW;  87D2100F979DF377 CRC64;

 Alignment Scores:
 Pred. No.:              0            Length:        977
 Score:                  4332.50      Matches:       820
 Percent Similarity:     89.86%       Conservative:  57
 Best Local Similarity:  84.02%       Mismatches:    90
 Query Match:            63.97%       Indels:        9
 DB:                     6            Gaps:          5

 US-09-302-812-5 (1-3814) x O02776 (1-977)

Qy        28 ATGAGTGCGGGCCCCGGCTGGGAGCCCTGCACGAAA---GCGCGCTGGGGCGCCGCTGGA 84
             ||||||||||||||||||   |||||||||||||||           ||||||   ||||||
Db         1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        85 ACTTCTGCGCCGACTGCCTCGGACTCCCGGAGCTTCCCTGGCAGGCAGAGGCGTGTTCTC 144
             ||||||   |||   |||||||||:::|||||||||||||||||||||||||||||||||
Db        21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy       145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGGTCCCTCCGTCCTCGCCAGCCTGCGTCTCG 204
             |||   ||||||||||||||||||||||||||||||||||||||||||       |||
Db        41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy       205 GGGCGGGCGGGACCGCACAGAGGCAACGCCACCTCGTTTGTTTTCAAACAAAAGACTATT 264
             ||||||||||||    |||||||||:::||||||||    |||||||||||||||||||||
Db        61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy       265 ACTACTTGGATGGATACTAAAGGACCCAAGACAGCTGAATCAGAA---------AGTAAA 315
             |||:::|||||||||||||||||||||||   ||||||   |||||||||     ||||||
Db        81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy       316 GAAAACAACAATACAAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTAC 375
             ||||||||||||||||||      :::|||||||||||||||||||||||||||||||||||
Db       101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy       376 CCACATAAGGTGGAAAAATTGGAAAATGTTCCTCAGCTAAATCTTGATAAATCACCCACA 435
                |||   :::||||||||||||||||||   ||||||       ||||||||||||
Db       121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy       436 GAAAAGAGTTCACAGTATTTGAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAAT 495
             ||||||    :::||||||||||   |||     ||||||||::::::||||||||||||||
Db       141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy       496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555
             ||||||   |||:::|||:::||||||    ||||||||||||||   |||    :::|||
Db       161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy       556 AAGCAGCTTAGTAATGCTAACATTGGTCAGTCACCCCACACTGATGACCACAGTGACACA 615
             :::|||   ||||||||||||:::     ||||||         ||||||||||||||||||
Db       181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy       616 GATCATGAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACA 675
             :::   ||||||   ||||||||||||||||||||||||||    :::||||||||||||
Db       201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy       676 AAGCCAACAGTAGGAGATGGGCAG---------GCCAGAAGCAACTGTAAGTGCAGTGGA 726
             |||   |||:::   |||   |||           |||||||||:::   ||||||
Db       221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy       727 TCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786
             :::         :::   |||   ||||||||||||||||||   ||||||:::
Db       241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy       787 GAGAGTCCTTTGTCAGATGTTGGTGCCGAGGACATTGGAACTGGACCAAAAAATGACAAC 846
             ||||||||||||||||||   |||:::||||||:::||||||||||   ||||||   |||
Db       261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy       847 AAATTGACTGGACAAGAAAGCAGCCTAGGTGATTCGCCTCCATTTGAGAAAGAAAGTGAG 906
             :::|||      ||||||||||||||||||:::|||||||||||||||||||||||||||||
Db       281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy       907 CCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGAT 966
             |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db       301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy       967 GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCCTCC---CAAACAGCAAAT 1023
             ||||||||||||||     ||||||||||||:::|||   ||||||     |||||||||||
Db       321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy      1024 AAACTTTCAAGTTGCCAAGCAAGAGAAGCTGATGGCGATCTTAGGAAACGGTATTTGACT 1083
             |||   |||         |||   ||||||||||||     :::||||||||||||
Db       341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy      1084 AAGGGAAGTGAAGTTAGATTGCATTTCCAATTTGAA---GGAGAAAATAATGCTGGGACC 1140
             ||||||   |||:::||||||||||||||||||||||     ||||||:::     ||||||
Db       361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy      1141 AGTGACTTAAATGCCAAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200
             :::|||:::|||||||||      |||::::::||||||||||||||||||||||||:::
Db       381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy      1201 TCCAAGCAGCATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAG 1260
             ||||||||||||||||||::::::|||||||||||||||||||||||||||||:::   |||
Db       401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy      1261 TCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATT 1320
             :::||||||:::|||||||||||||||||||:::::::||||||||||||||||||||||
Db       421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy      1321 CCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAG 1380
             ||||||||||||||||||:::|||   |||:::||||||||||||||||||||||||||||
Db       441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy      1381 GAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGT 1440
             ||||||||:::||||||||||||||||||   ||||||   |||||||||||||||:::
Db       461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy      1441 CACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
             ||||||||||||:::||||||||||||||||||   |||||||||||||||||||||||||
Db       481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy      1501 ACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTTGTTCGGAACAAAAC 1560
```

```
                  ||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy       1561 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCATT 1620
              |||||||||||||||||||||||||||||||||||||      ||||||||||||||||||||||
Db        521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy       1621 CAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAA 1680
              ||||||||||||||||:::    ||||||||||||||||||||||||||||||||||||
Db        541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy       1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAG 1740
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy       1741 GTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy       1801 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy       1861 AATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTC 1920
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       1921 TGCACATTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       1981 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACACGCTTACAC 2160
              ||||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2161 GTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2221 AATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTT 2280
              ||||||||||||||||||||||||||       ||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2281 TTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2401 CGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACG 2460
              ||||||||||||||||||||||     ||||||:::||||||||||||||||    |||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2521 AAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCT 2580
              |||:::||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2581 GAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2641 AGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTAT 2700
              ||||||||||||||||||||||||||||||||||||||||     ||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2701 TTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2761 GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAA 2820
              |||||||||||     ||||||:::|||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2821 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAAGGCAGGC 2928
              |||||||||      :::||||||:::     ||||||||:::    |||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 9
O46043
```
ID   O46043      PRELIMINARY;      PRT;   768 AA.
AC   O46043;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   PARG protein.
GN   PARG OR EG:114E2.1 OR CG2864.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ame J.-C., Jacobson M.K.;
RT   "Isolation and characterization of the cDNA encoding Drosophila
RT   poly(ADP-ribose) glycohydrolase.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 46-768 FROM N.A.
RA   Murphy L., Harris D., Barrell B.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 46-768 FROM N.A.
RA   Benos P.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003428; AAF45886.1; -.
DR   EMBL; AF079556; AAC28734.1; -.
DR   EMBL; Z98254; CAB10913.1; -.
DR   FlyBase; FBgn0023216; Parg.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   768 AA;  86288 MW;  2E620C42E643FC04 CRC64;

Alignment Scores:
Pred. No.:              1.47e-81       Length:          768
Score:                  1074.50        Matches:         244
Percent Similarity:     58.12%         Conservative:    96
Best Local Similarity:  41.71%         Mismatches:      194
Query Match:            15.86%         Indels:          51
DB:                     5              Gaps:            15

US-09-302-812-5 (1-3814) x O46043 (1-768)

Qy   1249 AGAATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACA 1308
Db     45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy   1309 GAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCA------GAGAAGAAGTGG 1362
Db     63 -----------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy   1363 CTGGGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCC 1422
Db     75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93

Qy   1423 TTAAGACCA---TCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGA 1479
Db     94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113

Qy   1480 GAGGTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAA 1539
Db    114 ThrProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisValArg 131

Qy   1540 ATGCCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCA 1599
Db    132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151

Qy   1600 GGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAAC 1659
Db    152 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171

Qy   1660 TTGAAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1719
Db    172 LeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191

Qy   1720 TTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCC 1779
Db    192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211

Qy   1780 ATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 1839
Db    212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 231

Qy   1840 CCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGT 1899
Db    232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 251

Qy   1900 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCC---AAGATGAAATCG 1956
Db    252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271

Qy   1957 GAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGG 2016
Db    272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291

Qy   2017 AAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTC------ACAGAGAAA 2070
Db    292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311

Qy   2071 AAA------------CCTACAGGATTGGTGACATTTACAAGACAGAGT------CTTGAA 2112
Db    312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331

Qy   2113 GATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTG------ACACGCTTACACGTCACT 2166
Db    332 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351

Qy   2167 TACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGT 2226
Db    352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371

Qy   2227 TTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATC 2286
Db    372 TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle 391

Qy   2287 AATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTT 2346
Db    392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411

Qy   2347 ATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG 2406
Db    412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp 431

Qy   2407 GCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATC 2466
Db    432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451

Qy   2467 GTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTG 2526
Db    452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471

Qy   2527 AGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT-----------------CCT 2568
Db    472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProPro 491

Qy   2569 GGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTT 2628
Db    492 Gly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503

Qy   2629 GGGGGTGACGCTAGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGT 2688
Db    504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523

Qy   2689 GACGTGGTTTATTTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC 2748
Db    524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrp 543

Qy   2749 ACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATAC 2808
Db    544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562
```

```
Qy       2809 TACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------GAC 2850
               |||:::         :::       |||:::      |||
Db        563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582

Qy       2851 ATCAAGCTTTATCCATTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGACATG 2910
                 |||||||||   ||||||                 |||      :::      |||:::
Db        583 LysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAspVal 599

Qy       2911 CCAGGACAGAAGGCA 2925
               ||||||:::   |||
Db        600 ProGlyGluGlyAla 604

RESULT 10
Q960N8
ID   Q960N8      PRELIMINARY;      PRT;   723 AA.
AC   Q960N8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   LD42380p.
GN   PARG OR EG:114E2.1 OR CG2864.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051955; AAK93379.1; -.
DR   FlyBase; FBgn0023216; Parg.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   723 AA;  81109 MW;  90A5BA59775D611C CRC64;

 Alignment Scores:
 Pred. No.:              3.79e-81      Length:         723
 Score:                  1069.50       Matches:        243
 Percent Similarity:     58.05%        Conservative:   96
 Best Local Similarity:  41.61%        Mismatches:     194
 Query Match:            15.79%        Indels:         51
 DB:                     5             Gaps:           15

US-09-302-812-5 (1-3814) x Q960N8 (1-723)

Qy       1252 ATTTCCAAGTCAGAGGACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAA 1311
                :::|||||||||   |||                :::    |||        :::
Db          1 MetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu--------- 17

Qy       1312 AGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCA------GAGAAGAAGTGGCTG 1365
                                     |||   ||||||             :::     |||
Db         18 ---------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrpArg 30

Qy       1366 GGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTA 1425
               |||      :::|||   :::   :::      |||            |||    :::|||
Db         31 GlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsnLeu 49

Qy       1426 AGACCA---TCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAG 1482
                 |||     :::|||    :::           |||          |||
Db         50 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 69

Qy       1483 GTTCCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATG 1542
                   |||:::|||:::    :::         |||       ||||||:::::||||||::::::
Db         70 ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisValArgLeu 87

Qy       1543 CCTTGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGG 1602
               ||||||:::    ::::::     ||||||     |||:::      :::|||        |||
Db         88 ProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIleAsp 107

Qy       1603 AGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTG 1662
                ||||||||:::|||:::       |||||||||                 :::   |||
Db        108 PheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGluLeu 127

Qy       1663 AAGGATGCGATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTG 1722
               :::     ||||||:::        ||||||        |||        :::|||  |||   ||||||
Db        128 GlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAlaLeu 147

Qy       1723 GTTGATTTCTGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATT 1782
                             |||:::   |||:::|||:::|||   :::      ::::::   :::
Db        148 HisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAspLeu 167

Qy       1783 TTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCA 1842
               ||||||   ::::::::::::||||||      ||||||::::::          |||:::|||
Db        168 LeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProValPro 187

Qy       1843 CTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAGATCGCCAGTCTT 1902
               |||||||||         |||      |||:::::::::|||||||:::||||||:::    |||
Db        188 LeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCysLeu 207

Qy       1903 TTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCC---AAGATGAAATCGGAG 1959
               |||||||||||||||   ||||||||||||||||||||||         |||    |||||||||
Db        208 LeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSerGlu 227

Qy       1960 TATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAA 2019
               ||||||::::::||||||||||||||||||||||||||||:::::::
Db        228 TyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAlaVal 247

Qy       2020 CCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTC------ACAGAGAAAAAA 2073
                 ||||||||||||      :::         |||||||||||||||        ||||||:::
Db        248 LeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArgAsp 267

Qy       2074 ------------CCTACAGGATTGGTGACATTTACAAGACAGAGT------CTTGAAGAT 2115
                           |||||||||:::|||||||||   |||:::|||            |||
Db        268 AlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGluHis 287

Qy       2116 TTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTG------ACACGCTTACACGTCACTTAC 2169
                  :::|||    :::            ||||||              |||||||||
Db        288 LeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAspAla 307

Qy       2170 GAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTT 2229
               |||||||||||||||         |||   |||:::|||||||||||||||||||||||||:::::
Db        308 GluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLysTyr 327

Qy       2230 GTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAAT 2289
               :::|||||||||||||      |||      |||      |||||||||||||||||||:::|||
Db        328 LeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIleCys 347

Qy       2290 CCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATT 2349
               |||||||||:::|||        :::||||||||||||||         |||       |||   |||:::
Db        348 ProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeuVal 367

Qy       2350 ATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCC 2409
                :::    |||     |||:::||||||      |||||||||||||||      ::::::   |||:::
Db        368 MetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrpSer 387

Qy       2410 CGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGGCGCTGCACGGAGATCGTT 2469
                  :::     ||||||    :::    :::|||                ||||||    |||    ||||||
Db        388 GlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIleVal 407

Qy       2470 GCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGA 2529
               ||||||||||||||||||      :::          |||:::       :::    :::
Db        408 AlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMetGlu 427

Qy       2530 CGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT------------------CCTGGA 2571
               ||||||||||||||||||||      ||||||:::                          ||||||
Db        428 ArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProProGly 447

Qy       2572 GTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGG 2631
                                         ||||||||||||||||||||||||||||||||||||
```

```
Db        448 -----------------------ValAlaThrGlyAsnTrpGlyCysGlyAlaPheGly 459

Qy       2632 GGTGACGCTAGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGAC 2691
Db        460 GlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArgPro 479

Qy       2692 GTGGTTTATTTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACT 2751
Db        480 LeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrpLeu 499

Qy       2752 TTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACTAC 2811
Db        500 LeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSerTyr 518

Qy       2812 AATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------GACATC 2853
Db        519 SerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLysLys 538

Qy       2854 AAGCTTTATCCATTCATATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGACATGCCA 2913
Db        539 LysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAspValPro 555

Qy       2914 GGACAGAAGGCA 2925
Db        556 GlyGluGlyAla 559
```

RESULT 11
Q9SKB3

```
ID   Q9SKB3      PRELIMINARY;      PRT;   548 AA.
AC   Q9SKB3; Q94ET7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative poly(ADP-ribose) glycohydrolase.
GN   AT2G31870 OR TEJ.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA   Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA   Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA   Fraser C.M., Venter J.C.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Panda S., Poirier G.G., Kay S.A.;
RT   "TEJ defines a role for poly-ADP-ribosylation in establishing period
RT   length of the Arabidopsis circadian oscillator.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006533; AAD32285.2; -.
DR   EMBL; AF394690; AAK72256.1; -.
DR   PIR; B84726; B84726.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   548 AA;  62169 MW;  F1A79FDA157C3329 CRC64;
```

Alignment Scores:

Pred. No.:	7.82e-39	Length:	548
Score:	568.00	Matches:	147
Percent Similarity:	46.12%	Conservative:	61
Best Local Similarity:	32.59%	Mismatches:	139
Query Match:	8.39%	Indels:	104
DB:	10	Gaps:	11

US-09-302-812-5 (1-3814) x Q9SKB3 (1-548)

```
Qy       1732 TGGGATAAGGTACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGAC 1791
Db         79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98

Qy       1792 ATGGTGAAAATTGCACTCTGTCTGCCAAATATT--------------------------- 1824
Db         99 LeuAlaSerLeuLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118

Qy       1825 ---------------TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCT 1869
Db        119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138

Qy       1870 GTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTT 1929
Db        139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158

Qy       1930 CCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAAT 1989
Db        159 ProAspAspAsnArgGlyAlaLys------------HisLeuProValIleAsnPheAsp 174

Qy       1990 CGGTTGTTTGAA------------GGACGTTCATCAAGGAAACCAGAAAAACTGAAAACA 2037
Db        175 HisLeuPheAlaSerLeuTyrIleSerTyrSerGlnSerGlnGluSerLysIleArgCys 194

Qy       2038 CTCTTCTGCTACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACA 2097
Db        195 IleMetHisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGlu 214

Qy       2098 AGACAG------AGTCTTGAAGATTTTCCAGAATGGGAAAGGTGTGAAAAGCCTCTGACA 2151
Db        215 ArgLysIleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeuCys 234

Qy       2152 CGCTTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTG 2211
Db        235 AlaPheLysValHisSerPheGlyLeuIleGluAspGlnProAspAsnAlaLeuGluVal 254

Qy       2212 GATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAA 2271
Db        255 AspPheAlaAsnLysTyrLeuGlyGlyGlySerLeuSerArgGlyCysValGlnGluGlu 274

Qy       2272 ATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGAT 2331
Db        275 IleArgPheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAsp 294

Qy       2332 CACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCT 2391
Db        295 AspAsnGluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAla 314

Qy       2392 GAAACTTATCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGGACGATTGGCAGCGG 2451
Db        315 SerSerPheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspProPheLysArg 334

Qy       2452 CGCTGCACGGAGATCGTTGCCATTGACGCACTT------------------CACTTCAGA 2493
Db        335 ArgArgThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLys 354

Qy       2494 CGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGC 2553
Db        355 AspIleCys---------------------LeuLeuArgGluIleAsnLysAlaLeuCys 367

Qy       2554 GGATTCCTC--------------------------------------------------- 2562
Db        368 GlyPheLeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAsp 387

Qy       2563 ------------------CGTCCTGGA--------------------------------- 2571
Db        388 AsnGluIleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThr 407

Qy       2572 ---------------------------------------------GTTCCTTCTGAAAAT 2586
Db        408 AlaSerHisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsn 427

Qy       2587 CTT---------------------------------------------TCTGCAGTG 2598
```

```
                |||                                                 |||
Db      428 LeuIleArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyVal 447

Qy     2599 GCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATA 2658
            ||||||||||||||||||||||||     |||||||||||||      |||||||||   |||
Db      448 AlaThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIle 467

Qy     2659 CAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTT---TATTTCACCTTTGGGGAC 2715
            |||   |||||||||:::              |||       |||:::|||||||||
Db      468 GlnTrpLeuAlaAlaSerGlnThrArgArgProPheIleSerTyrTyrThrPheGly--- 486

Qy     2716 TCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGGAGGAAGCTGGAT 2775
               |||    :::|||::::::       :::     ::::::             |||
Db      487 ValGluAlaLeuArgAsnLeuAspGlnValThrLysTrpIleLeuSerHisLysTrpThr 506

Qy     2776 GTTGGAAAAGTGTACAAGTTATTGCTTAGATAC 2808
            ||||||    ::::::   ::::::|||   |||
Db      507 ValGlyAspLeuTrpAsnMetMetLeuGluTyr 517
```

RESULT 12
Q8VYA1
```
ID   Q8VYA1      PRELIMINARY;      PRT;   522 AA.
AC   Q8VYA1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative poly(ADP-ribose) glycohydrolase.
GN   AT2G31870.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY072330; AAL61937.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   522 AA;  58942 MW;  6FF98685D867CA15 CRC64;
```

Alignment Scores:			
Pred. No.:	1.84e-35	Length:	522
Score:	528.00	Matches:	147
Percent Similarity:	46.10%	Conservative:	95
Best Local Similarity:	28.00%	Mismatches:	173
Query Match:	7.80%	Indels:	110
DB:	10	Gaps:	14

US-09-302-812-5 (1-3814) x Q8VYA1 (1-522)

```
Qy     1487 CGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCTT 1546
            |||::::::   |||::::::                                   |||
Db        8 ArgSerIleLeuGlnTyrLeuPro--------------------------------Leu 16

Qy     1547 GTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTA 1606
            |||          :::      :::|||          :::      |||||||||   :::
Db       17 ValAlaGlnSerSerSerLeuValTrpProProSerValGluGluGluLeuGlnThrIle 36

Qy     1607 -----------------------------GGTGGGAGCTCATTCAGACTGCACTTCTCA 1636
                                         |||    :::      ||||||   :::
Db       37 SerArgGlyProSerGluSerMetValAsnSerGlyGluAlaLeuAlaLeuHisIleThr 56

Qy     1637 ACAAATTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCATATT 1696
                :::      :::         |||         |||
Db       57 -AsnMetArgLysSerLeuSerLeuAsnAlaSerAspLeuAla----------------- 70

Qy     1697 CTAAGAAATGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAAGCAG 1756
                     |||      :::|||   |||:::|||||||   :::           |
Db       71 ----ProTyrAlaLeuGlnGlyTyrGly-LeuPhePheAspLysLysIleSerArgGluG 89

Qy     1757 AGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCACTCTGTCTGC 1816
            ||:::   :::   :::      ::::::|||   :::   ::::::   |||   ||||
Db       89 luSerAlaAsnPhePheGlyGluValValProAlaLeuCysArgLeuLeuLeuGlnLeuP 109

Qy     1817 CAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCAT---------- 1866
            ||::::::      :::                       ||||||   :::|||
Db      109 roSerMetLeuGluLysHisTyr------------GlnLysAlaAspHisValLeuAspG 125

Qy     1867 ----------------------------------------TCTGTCACGATGTCAC 1882
                                                        |||    :::||||
Db      125 lyValLysSerGlyLeuArgLeuLeuGlyProGlnGluAlaGlyIleValLeuLeuSerG 145

Qy     1883 AGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATG 1942
            |||||    ||||||:::|||||||||   :::|||||||||   ||||||      :::
Db      145 lnGluLeuIleAlaAlaLeuLeuAlaCysSerPhePheCysLeuPheProGluValAspA 165

Qy     1943 CCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAA- 2001
               :::|||        :::         |||||||||:::   ||||||
Db      165 rgSerLeuLys------------AsnLeuGlnGlyIleAsnPheSerGlyLeuPheSerP 181

Qy     2002 -----------GGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACT 2050
                          ::::::::         |||:::|||   |||      |||
Db      181 heProTyrMetArgHisCysThrLysGlnGluAsnLysIleLysCysLeuIleHisTyrP 201

Qy     2051 TTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTTACAAGACAGAGTCTTG 2110
             ||   |||:::            |||||||||   |||:::|||   |||:::   |||
Db      201 heGlyArgIleCysArgTrpMetProThrGlyPheValSerPheGluArgLysIleLeuP 221

Qy     2111 AA-----------------------GATTTTCCAGAA---------TGGGAAAGGTGTG 2137
                                        :::|||:::          |||
Db      221 roLeuGluTyrHisProHisPheValSerTyrProLysAlaAspSerTrpAlaAsnSerV 241

Qy     2138 AAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAG 2197
                ||||||      :::   :::                |||   ||||||
Db      241 alThrProLeuCysSerIleGluIleHisThrSerGlyAlaIleGluAspGlnProCysG 261

Qy     2198 GCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGAC 2257
               |||:::||||||||||||:::   :::   ||||||
Db      261 luAlaLeuGluValAspPheAlaAspGluTyrPheGlyGlyLeuThrLeuSerTyrAspT 281

Qy     2258 TTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCA 2317
               :::||||||||||||||||||:::||||||||||||||||||         :::|||
Db      281 hrLeuGlnGluGluIleArgPheValIleAsnProGluLeuIleAlaGlyMetIlePheL 301

Qy     2318 CTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAAT 2377
               :::|||   ||||||   :::   |||   |||   |||:::::::|||   |
Db      301 euProArgMetAspAlaAsnGluAlaIleGluIleValGlyValGluArgPheSerGlyT 321

Qy     2378 ACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATGGGAGTGAAAAGG 2437
              ||||||||||        ::::::::::::|||        :::   |||      :::   |
Db      321 yrThrGlyTyrGlyProSerPheGlnTyrAlaGlyAspTyrThrAspAsnLysAspLeuA 341

Qy     2438 ACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCT 2497
            ||   ::::::||||||   |||   ::::::||||||||||||:::
Db      341 spIlePheArgArgArgLysThrArgValIleAlaIleAspAlaMetPro---AspProG 360

Qy     2498 ACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGAT 2557
              :::   |||:::   :::   ||||||:::|||||||||:::      |||:
Db      360 lyMetGlyGlnTyrLysLeuAspAlaLeuIleArgGluValAsnLysAlaPheSerGlyT 380

Qy     2558 TCCTCCGT-----------------------------------------------CCTG 2569
              :::::
Db      380 yrMetHisGlnCysLysTyrAsnIleAspValLysHisAspProGluAlaSerSerSerH 400

Qy     2570 GAGTTCCTTCTGAAAATCTTTCTGCA-------------------------------- 2595
              ||||||      :::   ||||||
Db      400 isValProLeuThrSerAspSerAlaSerGlnValIleGluSerSerHisArgTrpCysI 420
```

```
Qy       2596 -------------------------GTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTG 2629
                                       ||||||||||||||||||||||||    ||||
Db        420 leAspHisGluGluLysLysIleGlyValAlaThrGlyAsnTrpGlyCysGlyValPheG 440

Qy       2630 GGGGTGACGCTAGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAACGTG 2689
              |||||||      ||||||   ::::::|||  |||||   :::   :::   |||
Db        440 lyGlyAspProGluLeuLysIleMetLeuGlnTrpLeuAlaIleSerGlnSerGlyArgP 460

Qy       2690 ACGTGGTT---TATTTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATGC 2746
                 :::    |||:::|||||||||      :::   ::::::::::::       :::
Db        460 roPheMetSerTyrTyrThrPheGly---LeuGlnAlaLeuGlnAsnLeuAsnGlnValI 479

Qy       2747 ACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGAT 2806
                 :::          :::::::::   ||||||   ::::::|||  |||:::    |
Db        479 leGluMetValAlaLeuGlnGluMetThrValGlyAspLeuTrpLysLysLeuValGluT 499

Qy       2807 ACTACAATGAA 2817
               ||   :::|||
Db        499 yrSerSerGlu 502
```

RESULT 13

```
Q9N5L4
ID   Q9N5L4      PRELIMINARY;      PRT;   485 AA.
AC   Q9N5L4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme
DE   4).
GN   H23L24.5 OR PME-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Abu-Threideh J., Lehnert L.;
RT   "The sequence of C. elegans cosmid H23L24.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Molecular characterization of poly ADP-ribose glycohydrolases in
RT   Caenorhabditis elegans.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006662; AAF39896.2; -.
DR   EMBL; AF548468; AAN40699.1; -.
DR   WormPep; H23L24.5; CE32685.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   485 AA;  55574 MW;  86367DDBAB15081F CRC64;

Alignment Scores:
Pred. No.:              1.46e-29      Length:         485
Score:                  458.00        Matches:        142
Percent Similarity:     46.69%        Conservative:   84
Best Local Similarity:  29.34%        Mismatches:     160

Query Match:            6.76%         Indels:         98
DB:                     5             Gaps:           18

US-09-302-812-5 (1-3814) x Q9N5L4 (1-485)

Qy       1609 TGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGA---CCCCAGAACTTGAAG 1665
              :::||||||:::::|||   :::      ::::::               ||||||||
Db         46 PheGluLeuLeuGluThrGlyValSerGlnGlnTrpArgAsnCysAspGlnAsnLeuPhe 65

Qy       1666 GATGCGATTCTGAAA------TACAATGTGGCATATTCT-------------------- 1698
              :::      ||||||       |||    ||||||
Db         66 AsnGluTyrLeuLysThrTyrLysAsnGlyGlyTyrSerGlnPheGluAspLeuLeuPhe 85

Qy       1699 --------------------------AAGAAATGGGACTTTACAGCTTTGGTTGATTTC 1731
                                        :::::::::|||        ||||||    |||
Db         86 LysIleTrpGlyTyrSerGluGluLysGluArgPheAspLeuProAlaLeuLysSerPhe 105

Qy       1732 TGGGATAAGGTACTTGAA---------GAAGCAGAGGCCCAACATTTATATCAGTCCATT 1782
              :::   |||:::   |||          |||  |||                    :::
Db        106 TyrArgLysMetSerGluIleValGlyGluAspGlu--------------------Val 118

Qy       1783 TTACCTGACATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCA 1842
              |||      :::   ::::::                 :::  |||
Db        119 LeuGluLysLeuAlaArgLeuValArgIleThrLysSerAlaCys-----------Glu 134

Qy       1843 CTCCTGAAACAGAAGATGAAT-----------------CATTCTGTCACGATGTCACAG 1884
              :::|||   :::|||:::                     |||   |||   |||
Db        135 ValLeuProGluLysIleTyrArgLeuValGlyAspIleGluSerAlaThrPheSerHis 154

Qy       1885 GAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCC 1944
                 |||   |||||||||:::|||      ||||||       ||||||
Db        155 IleGlnCysAlaSerLeuIleAlaTrpMetPhePheSerAspThrProArg--------- 171

Qy       1945 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGA 2004
                                                :::::|||    :::   :::
Db        172 --------------------------------LeuSerPheIleIleIleLeuGlnLys 180

Qy       2005 CGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACA 2064
                 :::           ||||||||||||   ||||||   ||||||   :::::::::
Db        181 ThrThrCysValAlaValGluLysLeuLysPheLeuPheThrTyrPheAspLysMetSer 200

Qy       2065 GAGAAAAAACCTACAGGATTGGTGACATTTACA-----------------AGACAGAGT 2106
                    |||   |||   |||:::|||                          :::|||
Db        201 IleAspProProIleGlyAlaValSerPheArgLysMetArgIleThrHisLysGlnTyr 220

Qy       2107 CTTGAAGATTTTCCAGAATGGGAA---AGGTGTGAAAAGCCTCTGACACGCTTACACGTC 2163
              ||||||:::              |||:::   |||               |||   :::   |||
Db        221 LeuGluAsn---------TrpLysLeuArgGluThrAsnLeuLeuProAspValGlnVal 237

Qy       2164 ACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAAT 2223
                    :::   :::||||||                 |||:::||||||||||||
Db        238 PheAspLysMetSerIleGluGluThrAlaLeu---CysThrGlnIleAspPheAlaAsn 256

Qy       2224 CGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTA 2283
              :::   :::||||||||||||          |||   ||||||||||||||||||:::
Db        257 LysArgLeuGlyGlyGlyValLeuLysGlyGlyAlaValGlnGluGluIleArgPheMet 276

Qy       2284 ATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGT 2343
              :::   ||||||::::::|||:::   |||      :::|||               |||
Db        277 MetCysProGluMetMetValAlaIleLeuLeuAsnAspValThrGlnAspLeuGluAla 296

Qy       2344 CTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGT 2403
              :::   |||   |||            :::|||   ||||||||||||:::   |||   :::
Db        297 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrSerAsnThrLeuLys 316

Qy       2404 TGGGCCCGA-----------------AGCCATGAAGATGGGAGTGAAAAGGACGATTGG 2445
              ||||||:::                        :::   :::   |||   :::|||   :::
Db        317 TrpAlaLysIleThrProLysHisSerAlaGlnAsnAsnAsnSerPheArgAspGlnPhe 336

Qy       2446 CAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC---------------TTC 2490
                 |||         ||||||   |||||||||||||||:::
Db        337 GlyArgLeuGlnThrGluThrValAlaIleAspAlaValArgAsnAlaGlyThrProLeu 356
```

```
Qy      2491 AGACGCTACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTAC 2550
                   |||::|||        ||||||:::  ||||||:::  ||||||
Db       357 GluCysLeuLeuAsnGlnLeuThrThrGluLysLeuThrArgGluValArgLysAlaAla 376

Qy      2551 TGCGGATTCCTCCGTCCTGGA---------------GTTCCTTCTGAAAATCTTTCTGCA 2595
             |||||||||       |||                 :::|||
Db       377 IleGlyPheLeuSerAlaGlyAspGlyPheSerLysIlePro------------------ 390

Qy      2596 GTGGCCACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTA 2655
             |||    :::|||  ||||||||||||||||   |||:::        ||||||   |||
Db       391 ValValSerGlyTrpTrpGlyCysGlyAlaPheArgGlyAsnLysProLeuLysPheLeu 410

Qy      2656 ATACAGATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGAC 2715
             ||||||:::::|||           ::::::|||   :::   :::   ||||||||||||
Db       411 IleGlnValIleAlaCysGlyIleSerAspArgProLeuGlnPheCysThrPheGlyAsp 430

Qy      2716 TCAGAGTTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGAT 2775
             :::||||||      :::            |||   |||                 :::
Db       431 ThrGluLeuAlaLysLysCysGluGluMetMetThrLeuPheArgAsnAsnAsnValArg 450

Qy      2776 GTTGGAAAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCC 2835
                |||:::::::::   ::::::                                   :::
Db       451 ThrGlyGlnLeuPheLeuIleIle--------------------------------Asn 459

Qy      2836 ACCCCTGGACCAGACATCAAGCTT---TATCCATTCATATACCATGCTGTTGAGTCAAGT 2892
             :::   ||||||   :::            :::::::::   |||:::   :::
Db       460 SerIleGlyProProLeuAsnTyrSerGluGlnTyrValPheAspAlaIleArgAlaLys 479

Qy      2893 GCAGAGACCACT 2904
                   :::|||
Db       480 IleAsnSerThr 483
```

RESULT 14

```
Q867X0
ID   Q867X0     PRELIMINARY;     PRT;   781 AA.
AC   Q867X0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3
DE   protein) (Corresponding sequence F20C5.1a).
GN   F20C5.1 OR PME-3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RT   Caenorhabditis elegans.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY185493; AAO26316.1; -.
DR   EMBL; Z68161; CAD89735.1; -.
DR   WormPep; F20C5.1a; CE33775.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   781 AA;  89273 MW;  30447D06605CED0A CRC64;

Alignment Scores:
Pred. No.:            6.81e-29      Length:          781
Score:                451.00        Matches:         197
Percent Similarity:          38.79%      Conservative:    137
Best Local Similarity:       22.88%      Mismatches:      297
Query Match:                 6.66%       Indels:          230
DB:                          5           Gaps:            37

US-09-302-812-5 (1-3814) x Q867X0 (1-781)

Qy       622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCA 681
             ::::::::::::|||   :::   ::::::         :::         |||:::   |||
Db        14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32

Qy       682 ACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741
             |||:::            :::   |||       :::          :::   :::|||
Db        33 ThrMet---------LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49

Qy       742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCA 801
             :::|||         :::   :::      ||||||         |||         |||
Db        50 GluAspProGlu------GluProLysSerArgAspValPheValSerSerGlnSerSer 67

Qy       802 GATGTTGGTGCCGAGGACATTGGAACTGGACCAAAA------------------------ 837
             |||         ||||||         |||:::
Db        68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87

Qy       838 ---------------------------------------------AATGACAACAAATTG 852
                                                             ||||||
Db        88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107

Qy       853 ACTGGACAAGAAAGCAGCCTAGGTGAT---------TCGCCTCCATTTGAGAAAGAAAGT 903
                   |||   :::|||   :::            :::   |||   |||::::::
Db       108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127

Qy       904 GAGCCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAGAA--- 960
                   :::|||   :::||||||         :::   ||||||   :::|||   |||
Db       128 AsnAsnLysSerAsnIleAspVal------AlaIleAsnSerAspGluAspAspGluLeu 145

Qy       961 ---------------------------------GCAGATGAAGAAACAAGTCCAGTCTTT 987
                                                  ::::::            :::|||
Db       146 ValLeuGluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnAspLeuPhe 165

Qy       988 ---GATGAGCAAGAT--------------------------------------------- 999
                |||:::|||:::
Db       166 AlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGlnLeu 185

Qy      1000 ---------GATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGAGAA 1050
                      |||         :::||||||   |||:::         :::      |||
Db       186 AspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThrGlu 205

Qy      1051 GCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCATTTC 1110
             ||||||
Db       206 AlaAspSer--------------------------------------------------- 208

Qy      1111 CAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGAAAC 1170
                |||   ||||||::::::                   |||   :::   ||||||
Db       209 ThrPheValGlyGluAspSerLys---------------AsnGlnArgGlnSerGlyThr 223

Qy      1171 TCT---------------------TCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAG 1209
             :::                      |||   :::|||:::      ::::::   :::
Db       224 ThrSerAspGluValAspAlaAspSerGlnIleAsnLeuAlaThrLysThrValArg--- 242

Qy      1210 CATGGAAAAAGGGATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAGGAC 1269
                            :::   :::               :::|||      |||
Db       243 ---------------ThrSerSerSerSerPheLeuSerThrValSerThrCysGluAla 257

Qy      1270 AGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGAACAGAAAGGAAGATTCCAAAATAC 1329
                |||   :::   :::   :::   |||:::   :::   :::
Db       258 ProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPhe 277

Qy      1330 ATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAAATGAGA 1389
                   ||||||
Db       278 ThrGluGlyAsnLeu--------------------------------------------- 282

Qy      1390 AAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTAAGACCATCTGCAAGTCACACAGTG 1449
                            :::|||:::|||          :::
```

```
Db     283 ----------------------------ThrLeuGlnProAspLeuAsn--------- 289

Qy    1450 ACTGTTCGGGTAGAC---------CTTCTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCA 1500
                 :::||||||              |||        :::|||     ||||||
Db     290 ------LysValAspProAspArgAsnTyrArgTyrCysThrIleProAsn---PhePro 306

Qy    1501 ACACATTACAAAGATTTG---TGGGATAACAAACAT---------GTGAAAATGCCTTGT 1548
                             |||      ||||||:::::            :::   :::|||
Db     307 AlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeuPro--- 325

Qy    1549 TCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAACTGCAGGGAGTAGG 1608
                                                                   |||
Db     326 ---------------------------------------------------GlnArg 327

Qy    1609 TGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACACGACCCCAGAACTTGAAGGAT 1668
           |||                                   :::|||            :::|||
Db     328 Trp-----------------------ArgGluPheAspSerArgGlyArgArgArgAsp 339

Qy    1669 GCGATTCTGAAATACAAT---------GTGGCATATTCTAAGAAATGGGACTTTACAGCT 1719
           :::         :::                     |||   |||    :::      |||
Db     340 SerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGly 359

Qy    1720 --------------------TTGGTTGATTTCTGGGAT---------AAGGTACTTGAA 1749
                               |||   :::   |||:::               :::
Db     360 TyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAspProAspIleThrTyr 379

Qy    1750 GAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTG------------ 1797
           :::         |||      :::|||       :::    ::::::|||
Db     380 LysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeuValGlyArgGluGlu 399

Qy    1798 --------------------------AAAATTGCACTCTGTCTGCCAAATATTTGCACC 1830
                                     :::||||||         :::|||
Db     400 ValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAlaGluAspIleLeuPro 419

Qy    1831 CAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCTGTCACGATGTCACAGGAACAG 1890
           :::     |||      |||          :::      |||  |||:::|||    :::|||
Db     420 GluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThrLeuSerHisLysGln 438

Qy    1891 ATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCCGACGGAATGCCAAGATG 1950
              |||:::|||:::|||         ||||||                      |||:::
Db     439 CysAlaAlaLeuValAlaArgMetPhePhe---------------------AlaArgPro 451

Qy    1951 AAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGGTTGTTTGAAGGACGTTCA 2010
              |||    :::|||                  |||    |||:::
Db     452 AspSerProPheSer-----------------PheCysArgIleLeuSerSerAspLys 465

Qy    2011 TCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGCTACTTTCGAAGAGTCACAGAGAAA 2070
           |||         ||||||||||||   ||||||   ||||||    :::::::::
Db     466 SerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAspLysMetSerMetAsp 485

Qy    2071 AAACCTACAGGATTGGTGACATTT------ACAAGACAGAGTCTTGAAGATTTT---CCA 2121
              |||   |||   |||:::|||      |||:::           :::   |||
Db     486 ProProAspGlyAlaValSerPheArgLeuThrLysMetAspLysAspThrPheAsnGlu 505

Qy    2122 GAATGG---GAAAGGTGTGAAAAGCCTCTGACACGCTTACACGTCACTTACGAGGGTACC 2178
           ||||||   ::::::      :::   |||      :::             |||
Db     506 GluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhePheAspGluMetLeu 525

Qy    2179 ATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGT 2238
           ||||||                  ||||||||||||||||||      :::||||||
Db     526 IleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsnGluHisLeuGlyGly 544

Qy    2239 GGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATCAGATTTTTAATCAATCCTGAATTG 2298
           ||||||         |||   |||||||||||||||||||||||:::   ||||||:::
Db     545 GlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeuMetCysProGluMet 564

Qy    2299 ATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCACAGGT 2358
           :::|||      |||      |||    :::            |||   :::   |||   |||
Db     565 MetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAlaIleSerIleValGly 584

Qy    2359 ACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCC--------- 2409
                 :::|||   ||||||||||||   |||   :::||||||
Db     585 AlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGln 604

Qy    2410 ------CGAAGCCATGAAGATGGGAGTGAA---AAGGACGATTGGCAGCGGCGCTGCACG 2460
                 |||   ::::::   :::|||   :::|||   :::   |||
Db     605 ProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPheGlyArgLeuArgVal 624

Qy    2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGC---------------TACCTCGAT 2505
           |||   :::||||||||||||:::   |||:::                        :::
Db     625 GluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGlu 644

Qy    2506 CAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTGCGGATTCCTCCGT 2565
           |||            :::   ||||||:::   ||||||         ||||||:::
Db     645 GlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSerIleGlyPheMetSer 664

Qy    2566 CCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTGCC 2625
              |||         |||:::      :::   ||||||   ||||||||||||||
Db     665 GlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAla 683

Qy    2626 TTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGCGGCTGAA 2685
           |||   ||||||      ||||||   :::||||||::::::||||||      |||:::
Db     684 PheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAsp 703

Qy    2686 CGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAGTTGATGAGAGACATTTACAGCATG 2745
           |||   :::   :::   :::||||||:::   ||||||                  :::
Db     704 ArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAlaLysCysLysLysIle 723

Qy    2746 CACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGA 2805
                 :::   ::::::   :::   :::|||   ::::::   ::::::
Db     724 IleGluArgMetLysGlnLysAspValThrLeuGlyMetLeuPheSerMetIle------ 741

Qy    2806 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCA 2865
                          |||         |||               :::   |||
Db     742 -----------------AsnAsnThrGlyLeuProHisLysHisPheGluPheTyrVal 755

Qy    2866 TTC 2868
           |||
Db     756 Phe 756
```

```
RESULT 15
Q19637
ID   Q19637     PRELIMINARY;     PRT;   764 AA.
AC   Q19637;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose
DE   metabolism enzyme-3 short form).
GN   F20C5.1 OR PME-3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RT   Caenorhabditis elegans.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z68161; CAA92299.2; -.
DR   EMBL; AY185494; AAO26317.1; -.
DR   PIR; T21138; T21138.
DR   WormPep; F20C5.1b; CE32867.
DR   InterPro; IPR007724; PARG.
```

```
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   764 AA;  87441 MW;  B391703B1D0CB79A CRC64;

Alignment Scores:
Pred. No.:              7.44e-29      Length:         764
Score:                  450.50        Matches:        187
Percent Similarity:     36.36%        Conservative:   129
Best Local Similarity:  21.52%        Mismatches:     290
Query Match:            6.65%         Indels:         263
DB:                     5             Gaps:           33

US-09-302-812-5 (1-3814) x Q19637 (1-764)

Qy    622 GAAGAAGACAGAGACAATCAGCAGTTTCTTACACCTATAAAACTTGCAAATACAAAGCCA 681
Db     14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32

Qy    682 ACAGTAGGAGATGGGCAGGCCAGAAGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTG 741
Db     33 ThrMet---------LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49

Qy    742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCAGAGAGTCCTTTGTCA 801
Db     50 GluAspProGlu------GluProLysSerArgAspValPheValSerSerGlnSerSer 67

Qy    802 GATGTTGGTGCCGAGGACATTGGAACTGGACCAAAA------------------------ 837
Db     68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87

Qy    838 ---------------------------------------------AATGACAACAAATTG 852
Db     88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107

Qy    853 ACTGGACAAGAAAGCAGCCTAGGTGAT---------TCGCCTCCATTTGAGAAAGAAAGT 903
Db    108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127

Qy    904 GAGCCTGAGTCACCAATGGATGTA------------------------------------ 927
Db    128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspAspGluLeuValLeu 147

Qy    928 ---GACAACTCGAGAAACAGTTGTCAAGATTCAGAAGCAGATGAAGAAACAAGTCCAGTC 984
Db    148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnLeuSerGlnAspLeu 167

Qy    985 TTT---GATGAGCAAGAT------------------------------------------ 999
Db    168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln 187

Qy   1000 ------------GATCGTTCCTCCCAAACAGCAAATAAACTTTCAAGTTGCCAAGCAAGA 1047
Db    188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207

Qy   1048 GAAGCTGATGGCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107
Db    208 GluAlaAspSer------------------------------------------------ 211

Qy   1108 TTCCAATTTGAAGGAGAAAATAATGCTGGGACCAGTGACTTAAATGCCAAGCCATCTGGA 1167
Db    212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSerSer 230

Qy   1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGAAAAAGG------ 1221
Db    231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249

Qy   1222 GATTCTAAAATTACAGATCATTTCATGAGAATTTCCAAGTCAGAG--------------- 1266
Db    250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269

Qy   1267 ---------------------GACAGAAGAAAAGAACAATGTGAAGTCAGACATCAAAGA 1305
Db    270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle------------ 285

Qy   1306 ACAGAAAGGAAGATTCCAAAATACATCCCACCTAACCTCCCTCCAGAGAAGAAGTGGCTG 1365
Db    286 ------------------------------ProAsnPheProAlaSerGln--------- 292

Qy   1366 GGAACTCCTATTGAGGAAATGAGAAAAATGCCTCGGTGTGGGATCCATTTGCCTTCCTTA 1425
Db    293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu--------------- 307

Qy   1426 AGACCATCTGCAAGTCACACAGTGACTGTTCGGGTAGACCTTCTGAGAGCAGGAGAGGTT 1485
Db    307 ------------------------------------------------------------ 307

Qy   1486 CCGAAACCTTTTCCAACACATTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCT 1545
Db    308 ------------ProGlnArgTrpArgGluPhe--------------------------- 314

Qy   1546 TGTTCGGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA------------ 1593
Db    315 ------------------------------AspSerArgGlyArgArgArgAspSerTyr 324

Qy   1594 ---------------------------------------------ACTGCAGGG------ 1602
Db    325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344

Qy   1603 ---------------------AGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAA 1641
Db    345 MetPheValGlyLeuLeuHisAsnMetTrpGluPhe------------------------ 356

Qy   1642 TTCACACGACCCCAGAACTTGAAGGATGCGATTCTGAAATACAATGTGGCA--------- 1692
Db    357 ------------------------AspProAspIleThrTyrLysLeuProAlaLeuGlu 368

Qy   1693 ------TATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTT 1746
Db    369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388

Qy   1747 GAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAAATTGCA 1806
Db    389 ArgValAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu------ArgIleTyr 406

Qy   1807 CTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCAT 1866
Db    407 ArgLeuValGlyAspVal---------------------------------------Glu 413

Qy   1867 TCTGTCACGATGTCACAGGAACAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACA 1926
Db    414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe------ 431

Qy   1927 TTTCCCCGACGGAATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTC 1986
Db    432 ---------------AlaArgProAspSerProPheSer------------------Phe 440

Qy   1987 AATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAAAAACTGAAAACACTCTTCTGC 2046
Db    441 CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr 460

Qy   2047 TACTTTCGAAGAGTCACAGAGAAAAAACCTACAGGATTGGTGACATTT------ACAAGA 2100
Db    461 TyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLys 480

Qy   2101 CAGAGTCTTGAAGATTTT---CCAGAATGG---GAAAGGTGTGAAAAGCCTCTGACACGC 2154
Db    481 MetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGlu 500

Qy   2155 TTACACGTCACTTACGAGGGTACCATAGAAGGCAACGGCCGAGGCATGCTACAGGTGGAT 2214
Db    501 ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp 519

Qy   2215 TTTGCAAATCGTTTTGTTGGAGGTGGTGTGACTGGTGCGGGACTTGTACAAGAAGAAATC 2274
Db    520 PheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIle 539

Qy   2275 AGATTTTTAATCAATCCTGAATTGATTGTTTCACGGCTGTTCACTGAGGTGCTGGATCAC 2334
```

```
              |||||||||:::  ||||||:::::|||      |||     |||  :::
Db       540 ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln 559

Qy      2335 AATGAGTGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAA 2394
                |||  :::  |||  |||         :::|||  ||||||||||||
Db       560 LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis 579

Qy      2395 ACTTATCGTTGGGCC---------------CGAAGCCATGAAGATGGGAGTGAA---AAG 2436
             |||  :::||||||                |||  ::::::  :::|||  :::
Db       580 ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg 599

Qy      2437 GACGATTGGCAGCGGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCACTTCAGACGC 2496
             |||  :::  |||          |||  :::||||||||||||:::  |||:::
Db       600 AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly 619

Qy      2497 ---------------TACCTCGATCAGTTTGTGCCTGAGAAAGTGAGACGTGAGCTTAAC 2541
                                 :::|||              :::  ||||||:::
Db       620 SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys 639

Qy      2542 AAGGCTTACTGCGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTGCAGTGGCC 2601
             ||||||      ||||||:::      |||         |||:::     :::
Db       640 LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal 658

Qy      2602 ACGGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAAGCCTTAATACAG 2661
             ||||||   ||||||||||||||||||    ||||||      ||||||   :::||||||
Db       659 ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln 678

Qy      2662 ATCCTGGCAGCTGCTGCGGCTGAACGTGACGTGGTTTATTTCACCTTTGGGGACTCAGAG 2721
             ::::::||||||      |||:::|||  :::  :::  :::||||||:::   |||
Db       679 ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu 698

Qy      2722 TTGATGAGAGACATTTACAGCATGCACACTTTCCTTACCGAGAGGAAGCTGGATGTTGGA 2781
             |||               :::       :::  ::::::  :::  :::|||
Db       699 LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly 718

Qy      2782 AAAGTGTACAAGTTATTGCTTAGATACTACAATGAAGAATGCAGAAACTGTTCCACCCCT 2841
                ::::::  ::::::                        |||         |||
Db       719 MetLeuPheSerMetIle-----------------------AsnAsnThrGlyLeuPro 730

Qy      2842 GGACCAGACATCAAGCTTTATCCATTC 2868
                       :::  |||  |||
Db       731 HisLysHisPheGluPheTyrValPhe 739
```

Search completed: May 26, 2004, 17:10:23
Job time : 268.048 secs

```
OM protein - protein search, using sw model

Run on:         May 26, 2004, 17:19:25 ; Search time 110.904 Seconds
                                         (without alignments)
                                         2466.140 Million cell updates/sec

Title:          US-09-302-812-6
Perfect score:  5164
Sequence:       1 MSAGPGWEPCTKARWGAAGT..........YHAVESSAETTDMPGQKAGT 968

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
               1:  geneseqp1980s:*
               2:  geneseqp1990s:*
               3:  geneseqp2000s:*
               4:  geneseqp2001s:*
               5:  geneseqp2002s:*
               6:  geneseqp2003as:*
               7:  geneseqp2003bs:*
               8:  geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5164	100.0	968	5	AAE25631	Aae25631 Murine po
2	5164	100.0	968	5	AAU76022	Aau76022 Mouse pol
3	5164	100.0	968	5	AAU76013	Aau76013 Mouse pol
4	5164	100.0	968	6	ABG72280	Abg72280 Mmurine p
5	4395	85.1	976	5	AAE25630	Aae25630 Human pol
6	4395	85.1	976	5	AAU76021	Aau76021 Human pol
7	4395	85.1	976	5	AAU76012	Aau76012 Human pol
8	4395	85.1	976	6	ABG72279	Abg72279 Human pol
9	4332.5	83.9	977	5	AAE25629	Aae25629 Bovine po
10	4332.5	83.9	977	5	AAU76020	Aau76020 Bovine po
11	4332.5	83.9	977	5	AAU75799	Aau75799 Bovine po
12	4332.5	83.9	977	6	ABG72278	Abg72278 Bovine po
13	1074.5	20.8	768	4	ABB59491	Abb59491 Drosophil
14	1074.5	20.8	768	5	AAE25632	Aae25632 Fruit fly
15	1074.5	20.8	768	5	AAU76023	Aau76023 Fruit fly
16	1074.5	20.8	768	5	AAU76014	Aau76014 Fruit fly
17	1074.5	20.8	768	6	ABG72281	Abg72281 Fruit fly
18	443.5	8.6	726	5	AAE25633	Aae25633 Poly aden
19	443.5	8.6	726	5	AAU76024	Aau76024 Worm poly
20	443.5	8.6	726	5	AAU76015	Aau76015 Worm poly
21	443.5	8.6	726	6	ABG72282	Abg72282 C. elegan
22	350.5	6.8	819	4	ABG20721	Abg20721 Novel hum
23	199	3.9	100	4	ABG11103	Abg11103 Novel hum
24	186	3.6	33	5	AAE25652	Aae25652 Bovine po
25	178.5	3.5	931	4	ABB61093	Abb61093 Drosophil
26	177	3.4	2703	4	ABB60074	Abb60074 Drosophil
27	172	3.3	1489	6	ABO14779	Abo14779 Novel hum
28	172	3.3	1489	6	ABU11038	Abu11038 Human pro
29	169.5	3.3	1297	6	ABU43768	Abu43768 Protein e
30	168	3.3	335	4	ABG09631	Abg09631 Novel hum
31	168	3.3	2951	4	ABB60291	Abb60291 Drosophil
32	167	3.2	31	5	AAE25634	Aae25634 Bovine po
33	167	3.2	31	5	AAE25651	Aae25651 Bovine po
34	167	3.2	31	5	AAU76025	Aau76025 Bovine po
35	167	3.2	31	5	AAU76016	Aau76016 Bovine po
36	167	3.2	31	6	ABG72283	Abg72283 Oligopept
37	167	3.2	1169	6	ABU43255	Abu43255 Protein e
38	167	3.2	1177	5	ABP40261	Abp40261 Staphyloc
39	165.5	3.2	787	6	ABP55414	Abp55414 Human MDD
40	165	3.2	1192	3	AAB18165	Aab18165 Plasmodiu
41	164.5	3.2	805	6	ABP55404	Abp55404 Human MDD
42	163.5	3.2	779	3	AAB42438	Aab42438 Human ORF
43	163.5	3.2	1187	2	AAR66451	Aar66451 AF-4 prot
44	163.5	3.2	1210	2	AAR66450	Aar66450 AF-4 prot
45	163.5	3.2	1210	7	ABR63878	Abr63878 Human AF4

ALIGNMENTS

```
RESULT 1
AAE25631
ID   AAE25631 standard; protein; 968 AA.
XX
AC   AAE25631;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Mus musculus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42083.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 63-68; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
```

```
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is murine PARG
XX
SQ   Sequence 968 AA;

  Query Match             100.0%;  Score 5164;  DB 5;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy        961 MPGQKAGT 968
              ||||||||
Db        961 MPGQKAGT 968

RESULT 2
AAU76022
ID   AAU76022 standard; protein; 968 AA.
XX
AC   AAU76022;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Mus musculus.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14933.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 63-70; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the mouse PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 968 AA;

  Query Match             100.0%;  Score 5164;  DB 5;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy   61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy  121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy  181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy  241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy  301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy  361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy  421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy  481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy  541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy  601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy  661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy  721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy  781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy  841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy  901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy  961 MPGQKAGT 968
        ||||||||
Db  961 MPGQKAGT 968
```

```
RESULT 3
AAU76013
ID   AAU76013 standard; protein; 968 AA.
XX
AC   AAU76013;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Mus musculus.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14495.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 63-68; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the mouse PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 968 AA;

  Query Match             100.0%;  Score 5164;  DB 5;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy   61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy  121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy  181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy  241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy        961 MPGQKAGT 968
              ||||||||
Db        961 MPGQKAGT 968

RESULT 4
ABG72280
ID   ABG72280 standard; protein; 968 AA.
XX
AC   ABG72280;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;
KW   cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;
KW   vasotropic; anticonvulsant; cerebroprotective; enzyme.
XX
OS   Mus musculus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14479.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents murine PARG enzyme
XX
SQ   Sequence 968 AA;

  Query Match             100.0%;  Score 5164;  DB 6;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy       421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy       481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy       541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy       601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy       661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy       721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy       781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy       841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy       901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy       961 MPGQKAGT 968
             ||||||||
Db       961 MPGQKAGT 968
```

RESULT 5
AAE25630
ID   AAE25630 standard; protein; 976 AA.
XX
AC   AAE25630;
XX
DT   04-NOV-2002 (first entry)
XX
DE   Human poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Homo sapiens.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42082.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 55-60; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is human PARG
XX
SQ   Sequence 976 AA;

```
  Query Match            85.1%;  Score 4395;  DB 5;  Length 976;
  Best Local Similarity  85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy         1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db         1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy        60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db        60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy       117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db       120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy       177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
Db       180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy       234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db       240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy       294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
Db       300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy       353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db       360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy       412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db       420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy       472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db       480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy       532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db       540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
```

```
Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
              ||||||| ||:|||||||||||||||| |||||||||||||||||:||||||||||||||
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              ||||||||||||||||||||:|||||||:|| :|||||||||||||||||||||||||||
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
              |:|||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
              |||||||||||||||| |||||||| || |||||||||||||||||||||||||||||||
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968
              ||| ||| |  ||: ||
Db        960 VESCAETADHSGQRTGT 976
```

RESULT 6
AAU76021

```
ID   AAU76021 standard; protein; 976 AA.
XX
AC   AAU76021;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Homo sapiens.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14932.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 55-60; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the human PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 976 AA;

  Query Match             85.1%;  Score 4395;  DB 5;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
              |:|||| ||||||| ||||  |:: |||:||| |||||||||| ||||||||||||||| 
Db          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
              |:|| |||:||| |||||||||:||||| ||||||   |||||||||:||||||||||||
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
              | |||| || ||:|||| |||| |||||| |||:: ||||||||| |||  ||   |  :|
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
              :| |||| :|| |||||| |:|||||||| :|||| | |  | |  ||  ||  |: |||
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
              |    :||  |||:|:|:|:||||||||:||:||| ||||||  ||| ||:|||||||||
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
              ||||||||||:||||||||||||||  |||:| |||||||| | |:||| ||  : |||  |
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
              || |||||||||| ||:  | :||||  ||  ||||||||:|||||:||||||| ||: |
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        412 SEDRRKEQCEVRHQTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
              :|||||| |  :|||||||||||:||:| |:|||||||||||:|||||| || |||||: 
Db        420 AEDRRKEQWETKHQTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
              ||||||| ||:|||||||||||||||| |||||||||||||||||:||||||||||||||
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              ||||||||||||||||||||:|||||||:|| :|||||||||||||||||||||||||||
```

```
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968

Db        960 VESCAETADHSGQRTGT 976
```

RESULT 7
AAU76012
ID   AAU76012 standard; protein; 976 AA.
XX
AC   AAU76012;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Homo sapiens.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999; 99US-00302812.
XX
PR   01-MAY-1998; 98US-0083768P.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14494.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 55-60; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the human PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 976 AA;

Query Match 85.1%; Score 4395; DB 5; Length 976;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

```
Qy          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Db          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968

Db        960 VESCAETADHSGQRTGT 976
```

RESULT 8
ABG72279
ID   ABG72279 standard; protein; 976 AA.

```
XX
AC   ABG72279;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14478.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents human PARG enzyme
XX
SQ   Sequence 976 AA;

  Query Match             85.1%;  Score 4395;  DB 6;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
               |:|||| |||||| ||||| |::| |||:|||| |||||||||| |||||||||||||
Db          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
              |:|| |||:||| |||||||:|||||| ||||||   ||||||||:||||||||||||
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
               | |||| || ||:|||| |||:||||| |||::||||||||| |||| |||     | :|
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
              :| ||||| :||  |||||| ||:||||||||| :|||| | |   |  ||    |: |||
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
              |   :||   |||:|:||:||:||||||:||:||| |||||  ||  |||:|||||||||
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
              ||||||||||:|||||||||||||| ||||:| |||||| |  |||:|| :  ||| |
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
              || |||||||| ||:   | :||||  || ||||||||:||||||:||||||| ||: |
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
              :|||||| | :|||||||||||:||:| |:|||||||||||:|||||| || |||||:
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
              |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
              |||||||| ||:|||||||||||||||||| |||||||||||||||:|||||||||||
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              ||||||||||||||||||||||:|||||||:|| :|||||||||||||||||||||||||
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
              |:||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
              |||||||||||||| ||||||| || |||||||||||||||||||||||||||||||||
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968
              ||| ||| |   ||: ||
Db        960 VESCAETADHSGQRTGT 976


RESULT 9
AAE25629
ID   AAE25629 standard; protein; 977 AA.
XX
AC   AAE25629;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).
XX
```

```
KW   Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Bos taurus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42081.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 47-45; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is bPARG
XX
SQ   Sequence 977 AA;

  Query Match             83.9%;  Score 4332.5;  DB 5;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
Db        961 VESCTQTTNQPGQRTG 976

RESULT 10
AAU76020
ID   AAU76020 standard; protein; 977 AA.
XX
AC   AAU76020;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Region          601. .617
FT                   /note= "Represents PARG oligopeptide #1"
FT   Region          761. .770
FT                   /note= "Represents PARG oligopeptide #2"
FT   Region          771. .801
```

```
FT                          /note= "Represents PARG oligopeptide #3"
FT   Region                 849. .880
FT                          /note= "Represents PARG oligopeptide #4"
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14931.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 47-52; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the bovine PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 977 AA;

  Query Match             83.9%;  Score 4332.5;  DB 5;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
Db        961 VESCTQTTNQPGQRTG 976

RESULT 11
AAU75799
ID   AAU75799 standard; protein; 977 AA.
XX
AC   AAU75799;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Bos taurus.
XX
FH   Key                    Location/Qualifiers
FT   Region                 601. .617
FT                          /note= "Represents PARG oligopeptide #1"
FT   Region                 761. .770
FT                          /note= "Represents PARG oligopeptide #2"
FT   Region                 771. .801
FT                          /note= "Represents PARG oligopeptide #3"
FT   Region                 849. .880
FT                          /note= "Represents PARG oligopeptide #4"
XX
PN   US6333148-B1.
```

```
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;    99US-00302812.
XX
PR   01-MAY-1998;    98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14493.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 45-52; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the bovine PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 977 AA;

  Query Match             83.9%;  Score 4332.5;  DB 5;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
Db        961 VESCTQTTNQPGQRTG 976


RESULT 12
ABG72278
ID   ABG72278 standard; protein; 977 AA.
XX
AC   ABG72278;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Bos taurus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;    98US-0083768P.
PR   30-APR-1999;    99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
```

```
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14477.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents bovine PARG enzyme
XX
SQ   Sequence 977 AA;

  Query Match             83.9%;  Score 4332.5;  DB 6;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
              |||||| |||||  || || || |  |||:||||||||||| ||||||||||||   |
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
              |||| |||:||| |||||||:|||||| || |||     ||||||| :|||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
               | :|||||| ||   ||||  || :|||  | |||::||||||| |:|:|| ||||| | :|
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
              :| ||||: ||    ||||| || |||||||| :|||| |  :|  |    |||: ||
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
               :     : | |||||| ||:  |||||  |:||:||| || |:|  ||||||:||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
              |||||||||:||||||||||||||| ||||:| || |||||| |  | ||||  :||||
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
              || |:||||||| ||: || :|:|||  |::||||||||:|||||::||||||||||: |
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
              :||:|||||||::||||||||||||||||:  ||:|||||||||||||:|||||| ||  |||||:
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
              ||||:||||||| |||||||||||:||||||||||||||||||||||||||| |||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
              |||||: ||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
              ||||||||||:||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              |||||||||||||||||||||||||||| ||:||||| ||||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
              |:||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
              |||||||||||||||||||||| ||:|||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
              |||  :||: |||: |
Db        961 VESCTQTTNQPGQRTG 976

RESULT 13
ABB59491
ID   ABB59491 standard; protein; 768 AA.
XX
AC   ABB59491;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5265.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03594.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
```

```
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1074.5;  DB 4;  Length 768;
  Best Local Similarity   41.7%;  Pred. No. 1.1e-80;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy        408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
              |:|||  |     :  |    :            | ||     :  |  |  :| : :  |     | :
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy        466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
              | |  :|  :    |     |         |:|: :   |   ||::||::||: :: || |: :|  |
Db         94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
                |||:|: |||         : |: ||: ||   |    :| |  ||      |: |:|:| :  ::
Db        152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
              :|| ::::|| ||::     |:||||    |  |:::||:||: ||||| |||||| | ||
Db        212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
              |||::||||||||::              |||| :   |||||  ||:         |||:|||  |:|   |
Db        272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy        696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
                 :| :     ||       |||  |||||   | |:||||||:::||||  |  | |||||||:|
Db        332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy        754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
                |||:| :||||  |     | |:: | |:|| |||| :: |: : || : :|      || |  |
Db        392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy        814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
              ||||||| :    |:  : : |||||| ||:            ||           ||||||||||
Db        452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy        868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
              |||: ||||:|::  |     |  :|:||: |  |   | :       |    |  ::: :||
Db        504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy        928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
              |:    :  |: |           ||| ||    |     :  |:||:  |
Db        563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604

RESULT 14
AAE25632
ID   AAE25632 standard; protein; 768 AA.
XX
AC   AAE25632;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy; fruit fly.
XX
OS   Drosophila melanogaster.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42084.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Example 13; Col 71-76; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is fruit fly PARG
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1074.5;  DB 5;  Length 768;
  Best Local Similarity   41.7%;  Pred. No. 1.1e-80;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy        408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
              |:|||  |     :  |    :            | ||     :  |  |  :| : :  |     | :
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy        466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
              | |  :|  :    |     |         |:|: :   |   ||::||::||: :: || |: :|  |
Db         94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
                |||:|: |||         : |: ||: ||   |    :| |  ||      |: |:|:| :  ::
Db        152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
              :|| ::::|| ||::     |:||||    |  |:::||:||: ||||| |||||| | ||
Db        212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
              |||::||||||||::              |||| :   |||||  ||:         |||:|||  |:|   |
Db        272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy        696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
                 :| :     ||       |||  |||||   | |:||||||:::||||  |  | |||||||:|
Db        332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy        754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
                |||:| :||||  |     | |:: | |:|| |||| :: |: : || : :|      || |  |
Db        392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy        814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
              ||||||| :    |:  : : |||||| ||:            ||           ||||||||||
Db        452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy        868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
```

```
               |||: ||||:|:: |     | : |:|||: |    | : |            | :::  :||
Db         504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy         928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
               |:   :  |:  |            ||| ||    |   :  |:||: |
Db         563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604


RESULT 15
AAU76023
ID   AAU76023 standard; protein; 768 AA.
XX
AC   AAU76023;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Fruit fly poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Fruit fly; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Drosophila melanogaster.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14934.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Example 13; Col 71-76; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the fruit fly PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1074.5;  DB 5;  Length 768;
  Best Local Similarity   41.7%;  Pred. No. 1.1e-80;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy         408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
               |:||| |     : |   :               | ||    :  | |  :| : :  |     | :
Db          45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy         466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
               | |  :| :      |    |          |:|: :  |  ||::||::||: :: || |: :| |
Db          94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy         525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
               |||:|: |||           : |: ||: ||   |  :| | ||     |: |:|:| :  ::
Db         152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy         585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
               :|| ::::|| ||::   |:||||   | |:::||:||: |||||| ||||||| | ||
Db         212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy         644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
               |||::|||||||||::         |||| :  |||||  ||:         |||:||| |:|   |
Db         272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy         696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
                :| :     ||          |||  |||||  | |:|||||||:::|||| |  | |||||||:|
Db         332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy         754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
               |||:| :||||  |    | |:::  | |:|| ||||| :: |: :  || :  :|    || |  |
Db         392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy         814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
               ||||||||  :     |:  :  : |||||| ||:              ||         ||||||||||
Db         452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy         868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
               |||: ||||:|:: |     | : |:|||: |    | : |            | :::  :||
Db         504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy         928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
               |:   :  |:  |            ||| ||    |   :  |:||: |
Db         563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604

Search completed: May 26, 2004, 18:40:07
Job time : 114.904 secs
```

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 31.0181 Seconds
(without alignments)
1611.120 Million cell updates/sec

Title: US-09-302-812-6
Perfect score: 5164
Sequence: 1 MSAGPGWEPCTKARWGAAGT..........YHAVESSAETTDMPGQKAGT 968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5164	100.0	968	4	US-09-302-812-6	Sequence 6, Appli
2	5164	100.0	968	4	US-09-511-477-6	Sequence 6, Appli
3	5164	100.0	968	4	US-09-511-507-6	Sequence 6, Appli
4	4395	85.1	976	4	US-09-302-812-4	Sequence 4, Appli
5	4395	85.1	976	4	US-09-511-477-4	Sequence 4, Appli
6	4395	85.1	976	4	US-09-511-507-4	Sequence 4, Appli
7	4332.5	83.9	977	4	US-09-302-812-2	Sequence 2, Appli
8	4332.5	83.9	977	4	US-09-511-477-2	Sequence 2, Appli
9	4332.5	83.9	977	4	US-09-511-507-2	Sequence 2, Appli
10	1074.5	20.8	768	4	US-09-302-812-8	Sequence 8, Appli
11	1074.5	20.8	768	4	US-09-511-477-8	Sequence 8, Appli
12	1074.5	20.8	768	4	US-09-511-507-8	Sequence 8, Appli
13	443.5	8.6	726	4	US-09-302-812-10	Sequence 10, Appl
14	443.5	8.6	726	4	US-09-511-477-10	Sequence 10, Appl
15	443.5	8.6	726	4	US-09-511-507-10	Sequence 10, Appl
16	167	3.2	31	4	US-09-302-812-11	Sequence 11, Appl
17	167	3.2	31	4	US-09-511-477-11	Sequence 11, Appl
18	167	3.2	31	4	US-09-511-507-11	Sequence 11, Appl
19	167	3.2	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
20	163.5	3.2	1187	1	US-08-320-559-28	Sequence 28, Appl
21	163.5	3.2	1187	3	US-08-545-860D-28	Sequence 28, Appl
22	163.5	3.2	1187	5	PCT-US94-04496-28	Sequence 28, Appl
23	163.5	3.2	1210	1	US-08-320-559-26	Sequence 26, Appl
24	163.5	3.2	1210	3	US-08-545-860D-26	Sequence 26, Appl
25	163.5	3.2	1210	5	PCT-US94-04496-26	Sequence 26, Appl
26	153	3.0	1792	2	US-08-962-284-4	Sequence 4, Appli
27	150	2.9	1115	2	US-08-568-459A-2	Sequence 2, Appli
28	150	2.9	1115	2	US-08-487-826B-2	Sequence 2, Appli
29	150	2.9	1115	4	US-09-210-288-2	Sequence 2, Appli
30	150	2.9	1115	6	5198347-6	Patent No. 5198347
31	145.5	2.8	559	1	US-08-320-559-31	Sequence 31, Appl
32	145.5	2.8	559	3	US-08-545-860D-31	Sequence 31, Appl
33	145.5	2.8	559	5	PCT-US94-04496-31	Sequence 31, Appl
34	144	2.8	26	4	US-09-302-812-21	Sequence 21, Appl
35	144	2.8	26	4	US-09-511-477-21	Sequence 21, Appl
36	144	2.8	26	4	US-09-511-507-21	Sequence 21, Appl
37	144	2.8	29	4	US-09-302-812-12	Sequence 12, Appl
38	144	2.8	29	4	US-09-511-477-12	Sequence 12, Appl
39	144	2.8	29	4	US-09-511-507-12	Sequence 12, Appl
40	137.5	2.7	913	4	US-08-971-089-4	Sequence 4, Appli
41	137	2.7	1805	1	US-07-853-913-2	Sequence 2, Appli
42	136.5	2.6	1435	2	US-08-568-459A-4	Sequence 4, Appli
43	136.5	2.6	1435	2	US-08-487-826B-4	Sequence 4, Appli
44	136.5	2.6	1435	4	US-09-210-288-4	Sequence 4, Appli
45	136.5	2.6	1848	3	US-08-296-791-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

```
  Query Match             100.0%;  Score 5164;  DB 4;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
```

```
Db    301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy    961 MPGQKAGT 968
          ||||||||
Db    961 MPGQKAGT 968
```

```
RESULT 2
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-511-477-6
```

```
  Query Match            100.0%;  Score 5164;  DB 4;  Length 968;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  968;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy     61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy    121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy    181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy    241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy    301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy    961 MPGQKAGT 968
          ||||||||
Db    961 MPGQKAGT 968
```

```
RESULT 3
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-511-507-6

  Query Match             100.0%;  Score 5164;  DB 4;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKD 240

Qy        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy        961 MPGQKAGT 968
              ||||||||
Db        961 MPGQKAGT 968

RESULT 4
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-302-812-4

  Query Match             85.1%;  Score 4395;  DB 4;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
              |:|||| |||||| ||||| |::| |||:|||| ||||||||||| ||||||||||||| 
Db          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
              |:|| |||:||| ||||||||:|||||| ||||||   |||||||||:||||||||||||
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
               | |||| || ||:|||| ||||:||||| |||::||||||||| |||| |||   | :|
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
              :| |||||  :||  ||||||| ||:|||||||| :|||| | | | ||   |: ||| 
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
              |    :||  |||:|:||:|:|||||||:||:||| ||||||  ||| ||:|||||||||
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
```

```
Qy   294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
Db   300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy   353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db   360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy   412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db   420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy   472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db   480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy   532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db   540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy   592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db   600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy   652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db   660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy   712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db   720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy   772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db   780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy   832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db   840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy   892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db   900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy   952 VESSAETTDMPGQKAGT 968
Db   960 VESCAETADHSGQRTGT 976
```

```
RESULT 5
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
US-09-511-477-4

  Query Match             85.1%;  Score 4395;  DB 4;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy     1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db     1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy    60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db    60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy   117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db   120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy   177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
Db   180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy   234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db   240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy   294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
Db   300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy   353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db   360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy   412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db   420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy   472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db   480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy   532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db   540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy   592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db   600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy   652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db   660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy   712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db   720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy   772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db   780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy   832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db   840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy   892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db   900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy   952 VESSAETTDMPGQKAGT 968
Db   960 VESCAETADHSGQRTGT 976
```

```
RESULT 6
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-507-4

  Query Match             85.1%;  Score 4395;  DB 4;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
Db        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
Db        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968
Db        960 VESCAETADHSGQRTGT 976


RESULT 7
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-302-812-2

  Query Match             83.9%;  Score 4332.5;  DB 4;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
```

```
Db     241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy     294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db     301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy     353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db     361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy     412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db     421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy     472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db     481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy     532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db     541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy     592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db     601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy     652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db     661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy     712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db     721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy     772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db     781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy     832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db     841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy     892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db     901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy     952 VESSAETTDMPGQKAG 967
Db     961 VESCTQTTNQPGQRTG 976
```

RESULT 8
US-09-511-477-2
; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-2

```
  Query Match             83.9%;  Score 4332.5;  DB 4;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy       1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db       1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy      60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db      61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy     117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db     121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy     177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db     181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy     234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db     241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy     294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db     301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy     353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db     361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy     412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db     421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy     472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db     481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy     532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db     541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy     592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db     601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy     652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db     661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy     712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db     721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy     772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db     781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy     832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db     841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy     892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db     901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy     952 VESSAETTDMPGQKAG 967
```

```
Db        961 VESCTQTTNQPGQRTG 976

RESULT 9
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-511-507-2

  Query Match            83.9%;  Score 4332.5;  DB 4;  Length 977;
  Best Local Similarity  84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
Db        961 VESCTQTTNQPGQRTG 976

RESULT 10
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-302-812-8

  Query Match            20.8%;  Score 1074.5;  DB 4;  Length 768;
  Best Local Similarity  41.7%;  Pred. No. 6.9e-96;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy        408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy        466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
Db         94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
Db        152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
Db        212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271
```

```
Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
              |||::||||||||::          ||||  :  |||||  ||:       |||:|||  |:|     |
Db        272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy        696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
                :|  :      ||       |||   |||||   | :|||||||:::||||  |  | |||||||:|
Db        332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy        754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
                |||:|  :||||  |    | |::  |  |:||  |||||  ::  |:  :  ||  :  :|      ||  |  |
Db        392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy        814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
              |||||||  :      |:   :  :  ||||||  ||:              ||             ||||||||||
Db        452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy        868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
              |||:  ||||:|::   |      |  :  |:|||:  |       |  :  |                 |  :::   :||
Db        504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy        928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
              |:    :   |:    |              |||  ||         |      :    |:||:  |
Db        563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604
```

RESULT 11
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-511-477-8

```
  Query Match             20.8%;  Score 1074.5;  DB 4;  Length 768;
  Best Local Similarity   41.7%;  Pred. No. 6.9e-96;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy        408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
              |:|||  |          :  |       :                     |  ||          :        |  |    :|  :   :       |              |  :
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy        466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
              |  |  :|  :         |           |     |:|:  :        |  ||::||::||:  ::  ||  |:  :|    |
Db         94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
                 |||:|:  |||           :  |:  ||:  ||       |    :|   |   |  ||          |:  |:|:|   :    ::
Db        152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
               :||  ::::||  ||::      |:||||          |  |:::||:||:  ||||||  |||||||       |  ||
Db        212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
              |||::||||||||::          ||||  :  |||||  ||:       |||:|||  |:|     |
```

```
Db        272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy        696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
                :|  :      ||       |||   |||||   | :|||||||:::||||  |  | |||||||:|
Db        332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy        754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
                |||:|  :||||  |    | |::  |  |:||  |||||  ::  |:  :  ||  :  :|      ||  |  |
Db        392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy        814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
              |||||||  :      |:   :  :  ||||||  ||:              ||             ||||||||||
Db        452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy        868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
              |||:  ||||:|::   |      |  :  |:|||:  |       |  :  |                 |  :::   :||
Db        504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy        928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
              |:    :   |:    |              |||  ||         |      :    |:||:  |
Db        563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604
```

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-511-507-8

```
  Query Match             20.8%;  Score 1074.5;  DB 4;  Length 768;
  Best Local Similarity   41.7%;  Pred. No. 6.9e-96;
  Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy        408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
              |:|||  |          :  |       :                     |  ||          :        |  |    :|  :   :       |              |  :
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy        466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
              |  |  :|  :         |           |     |:|:  :        |  ||::||::||:  ::  ||  |:  :|    |
Db         94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
                 |||:|:  |||           :  |:  ||:  ||       |    :|   |   |  ||          |:  |:|:|   :    ::
Db        152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
               :||  ::::||  ||::      |:||||          |  |:::||:||:  ||||||  |||||||       |  ||
Db        212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
              |||::||||||||::          ||||  :  |||||  ||:       |||:|||  |:|     |
Db        272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy        696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
```

```
Db        332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy        754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813

Db        392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy        814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867

Db        452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy        868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927

Db        504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy        928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966

Db        563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604
```

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-302-812-10

```
  Query Match             8.6%;  Score 443.5;  DB 4;  Length 726;
  Best Local Similarity   21.9%;  Pred. No. 5.2e-34;
  Matches  184;  Conservative  124;  Mismatches  278;  Indels  255;  Gaps   32;

Qy        199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258

Db         14 QDEKDYEDYV-GVGFAHQVPTM---KRRKLTEHGNTTESKEDPE--EPKSRDVFVSSQSS 67

Qy        259 DVGAEDIGTGPK----------------------NDNKLTGQESSLGD---SPPFEKES 292

Db         68 DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV 127

Qy        293 EPESPMDV------------DNSRNSCQDSEADEETSPVF-DEQD------------- 324

Db        128 NNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIEYPGIMKDTTTQ 187

Qy        325 ----DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSG 380

Db        188 LDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSKATKTVRTSSSS 230

Qy        381 NSSSLNVECRSSKQHGKR--DSKITDHFMRISKSE-----------DRRKEQCEVRHQR 426

Db        231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLQPDLNKVDPDRNYRYCTI---- 285

Qy        427 TERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486

Db        286 ----------PNFPASQ---GKLREDNRYGPKIVL------------------------ 307

Qy        487 PKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER------------------TAG-- 525

Db        308 ----PQRWREF-------------------DSRGRRRDSYFYFKRKLDGYLKCYKTTGYF 344

Qy        526 -------SRWELIQTALLNKFTRPQNLKDAILKYNVA-----YSKKWDFTALVDFWDKVL 573

Db        345 MFVGLLHNMWEF----------------DPDITYKLPALEMYYKEMSELVGREEVLEKFA 388

Qy        574 EEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT 633

Db        389 RVARIAKTAEDILPE--RIYRLVGDV-------------ESATLSHKQCAALVARMFF-- 431

Qy        634 FPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TR 691

Db        432 -----ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTK 480

Qy        692 QSLEDF-PEW-ERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEI 749

Db        481 MDKDTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEI 539

Qy        750 RFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDGSE-K 803

Db        540 RFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFR 599

Qy        804 DDWQRRCTEIVAIDALHFRR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSENLSAVA 858

Db        600 DRFGRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIV 658

Qy        859 TGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVG 918

Db        659 TGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLG 718

Qy        919 K 919

Db        719 K 719
```

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-477-10

```
  Query Match             8.6%;  Score 443.5;  DB 4;  Length 726;
  Best Local Similarity   21.9%;  Pred. No. 5.2e-34;
  Matches  184;  Conservative  124;  Mismatches  278;  Indels  255;  Gaps   32;

Qy        199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258

Db         14 QDEKDYEDYV-GVGFAHQVPTM---KRRKLTEHGNTTESKEDPE--EPKSRDVFVSSQSS 67

Qy        259 DVGAEDIGTGPK----------------------NDNKLTGQESSLGD---SPPFEKES 292

Db         68 DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV 127
```

```
Qy     293 EPESPMDV-------------DNSRNSCQDSEADEETSPVF-DEQD-------------- 324
Db     128 NNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIEYPGIMKDTTTQ 187

Qy     325 ----DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSG 380
Db     188 LDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSKATKTVRTSSSS 230

Qy     381 NSSSLNVECRSSKQHGKR--DSKITDHFMRISKSE------------DRRKEQCEVRHQR 426
Db     231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLQPDLNKVDPDRNYRYCTI---- 285

Qy     427 TERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486
Db     286 ----------PNFPASQ---GKLREDNRYGPKIVL------------------------- 307

Qy     487 PKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER-------------------TAG-- 525
Db     308 ----PQRWREF------------------DSRGRRRDSYFYFKRKLDGYLKCYKTTGYF 344

Qy     526 -------SRWELIQTALLNKFTRPQNLKDAILKYNVA-----YSKKWDFTALVDFWDKVL 573
Db     345 MFVGLLHNMWEF----------------DPDITYKLPALEMYYKEMSELVGREEVLEKFA 388

Qy     574 EEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT 633
Db     389 RVARIAKTAEDILPE--RIYRLVGDV-------------ESATLSHKQCAALVARMFF-- 431

Qy     634 FPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TR 691
Db     432 -----ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTK 480

Qy     692 QSLEDF-PEW-ERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEI 749
Db     481 MDKDTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEI 539

Qy     750 RFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDGSE-K 803
Db     540 RFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFR 599

Qy     804 DDWQRRCTEIVAIDALHFRR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSENLSAVA 858
Db     600 DRFGRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIV 658

Qy     859 TGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVG 918
Db     659 TGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLG 718

Qy     919 K 919
Db     719 K 719

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-507-10

  Query Match             8.6%;  Score 443.5;  DB 4;  Length 726;
  Best Local Similarity   21.9%;  Pred. No. 5.2e-34;
  Matches  184;  Conservative  124;  Mismatches  278;  Indels  255;  Gaps   32;

Qy     199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258
Db      14 QDEKDYEDYV-GVGFAHQVPTM---KRRKLTEHGNTTESKEDPE--EPKSRDVFVSSQSS 67

Qy     259 DVGAEDIGTGPK----------------------NDNKLTGQESSLGD---SPPFEKES 292
Db      68 DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV 127

Qy     293 EPESPMDV-------------DNSRNSCQDSEADEETSPVF-DEQD-------------- 324
Db     128 NNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIEYPGIMKDTTTQ 187

Qy     325 ----DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSG 380
Db     188 LDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSKATKTVRTSSSS 230

Qy     381 NSSSLNVECRSSKQHGKR--DSKITDHFMRISKSE------------DRRKEQCEVRHQR 426
Db     231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLQPDLNKVDPDRNYRYCTI---- 285

Qy     427 TERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486
Db     286 ----------PNFPASQ---GKLREDNRYGPKIVL------------------------- 307

Qy     487 PKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER-------------------TAG-- 525
Db     308 ----PQRWREF------------------DSRGRRRDSYFYFKRKLDGYLKCYKTTGYF 344

Qy     526 -------SRWELIQTALLNKFTRPQNLKDAILKYNVA-----YSKKWDFTALVDFWDKVL 573
Db     345 MFVGLLHNMWEF----------------DPDITYKLPALEMYYKEMSELVGREEVLEKFA 388

Qy     574 EEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT 633
Db     389 RVARIAKTAEDILPE--RIYRLVGDV-------------ESATLSHKQCAALVARMFF-- 431

Qy     634 FPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TR 691
Db     432 -----ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTK 480

Qy     692 QSLEDF-PEW-ERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEI 749
Db     481 MDKDTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEI 539

Qy     750 RFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDGSE-K 803
Db     540 RFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFR 599

Qy     804 DDWQRRCTEIVAIDALHFRR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSENLSAVA 858
Db     600 DRFGRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIV 658

Qy     859 TGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVG 918
Db     659 TGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLG 718

Qy     919 K 919
Db     719 K 719

Search completed: May 26, 2004, 18:49:30
Job time : 35.0181 secs
```

```
                     GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 26, 2004, 18:46:21 ; Search time 80.7642 Seconds
                                          (without alignments)
                                          3343.018 Million cell updates/sec

Title:           US-09-302-812-6
Perfect score:   5164
Sequence:        1 MSAGPGWEPCTKARWGAAGT..........YHAVESSAETTDMPGQKAGT 968

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters:      1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                 8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                 10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                 12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                 13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                 14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                 15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                 16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                 17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                 18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5164	100.0	968	9	US-09-973-451-6	Sequence 6, Appli
2	4395	85.1	976	9	US-09-973-451-4	Sequence 4, Appli
3	4332.5	83.9	977	9	US-09-973-451-2	Sequence 2, Appli
4	1074.5	20.8	768	9	US-09-973-451-8	Sequence 8, Appli
5	579	11.2	546	12	US-10-425-114-60000	Sequence 60000, A
6	443.5	8.6	726	9	US-09-973-451-10	Sequence 10, Appl
7	220.5	4.3	200	12	US-10-424-599-184988	Sequence 184988,
8	176	3.4	6642	15	US-10-369-493-5013	Sequence 5013, Ap
9	175	3.4	180	12	US-10-424-599-156445	Sequence 156445,
10	172.5	3.3	300	12	US-10-424-599-233915	Sequence 233915,
11	172	3.3	1489	15	US-10-115-482-18	Sequence 18, Appl
12	169.5	3.3	1297	12	US-10-282-122A-71692	Sequence 71692, A
13	167	3.2	31	9	US-09-973-451-11	Sequence 11, Appl
14	167	3.2	1169	12	US-10-282-122A-71179	Sequence 71179, A
15	163.5	3.2	1210	14	US-10-205-823-264	Sequence 264, App
16	161	3.1	748	9	US-09-864-761-43244	Sequence 43244, A
17	159.5	3.1	1501	9	US-09-924-154-17	Sequence 17, Appl
18	159	3.1	1462	14	US-10-287-218-17	Sequence 17, Appl
19	159	3.1	2677	15	US-10-144-194A-22	Sequence 22, Appl
20	159	3.1	2697	15	US-10-144-198-12	Sequence 12, Appl
21	158.5	3.1	4723	15	US-10-359-012-8	Sequence 8, Appli
22	156	3.0	691	15	US-10-108-260A-4345	Sequence 4345, Ap
23	150	2.9	447	10	US-09-794-422-44	Sequence 44, Appl
24	150	2.9	1115	13	US-10-153-273-2	Sequence 2, Appli
25	148	2.9	1389	12	US-10-282-122A-70151	Sequence 70151, A
26	147.5	2.9	1358	15	US-10-369-493-1658	Sequence 1658, Ap
27	147.5	2.9	1973	16	US-10-197-824-34	Sequence 34, Appl
28	147.5	2.9	2047	12	US-10-346-863-2	Sequence 2, Appli
29	147.5	2.9	3863	16	US-10-197-824-7	Sequence 7, Appli
30	145.5	2.8	1881	14	US-10-032-585-7646	Sequence 7646, Ap
31	144.5	2.8	747	12	US-10-425-114-67803	Sequence 67803, A
32	144	2.8	26	9	US-09-973-451-21	Sequence 21, Appl
33	144	2.8	29	9	US-09-973-451-12	Sequence 12, Appl
34	143	2.8	1253	14	US-10-363-798-2	Sequence 2, Appli
35	143	2.8	2158	15	US-10-341-434-222	Sequence 222, App
36	143	2.8	2158	15	US-10-144-198-34	Sequence 34, Appl
37	143	2.8	2265	15	US-10-144-198-35	Sequence 35, Appl
38	142	2.7	1237	15	US-10-108-260A-3795	Sequence 3795, Ap
39	141.5	2.7	677	14	US-10-060-036-180	Sequence 180, App
40	140.5	2.7	3067	10	US-09-949-029-18	Sequence 18, Appl
41	140	2.7	1309	15	US-10-369-493-1656	Sequence 1656, Ap
42	140	2.7	1325	9	US-09-864-761-35612	Sequence 35612, A
43	140	2.7	1460	15	US-10-295-027-428	Sequence 428, App
44	139.5	2.7	498	12	US-10-424-599-196154	Sequence 196154,
45	139.5	2.7	791	15	US-10-144-194A-74	Sequence 74, Appl

ALIGNMENTS

```
RESULT 1
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-973-451-6

  Query Match             100.0%;  Score 5164;  DB 9;  Length 968;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  968;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSG 60

Qy         61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 RAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKV 120

Qy    121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 EKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLS 180

Qy    181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSGRQSVKD 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 NANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSGRQSVKD 240

Qy    241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300

Qy    301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLH 360

Qy    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 FQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC 420

Qy    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 EVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDL 480

Qy    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 LRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFT 540

Qy    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 RPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNIC 600

Qy    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGR 660

Qy    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 SSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEG 720

Qy    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 NGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQ 780

Qy    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 YSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKA 840

Qy    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 YCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900

Qy    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 RDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAVESSAETTD 960

Qy    961 MPGQKAGT 968
          ||||||||
Db    961 MPGQKAGT 968
```

RESULT 2
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI\
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-973-451-4

```
  Query Match             85.1%;  Score 4395;  DB 9;  Length 976;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy      1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
          |:|||| |||||| ||||| |::| |||:|||| ||||||||||| |||||||||||||
Db      1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59

Qy     60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
          |:|| |||:||| ||||||||:|||||| ||||||   |||||||||:||||||||||||
Db     60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119

Qy    117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
           | |||| || ||:|||| ||||:||||| |||::||||||||| |||| |||   | :|
Db    120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179

Qy    177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233
          :| ||||| :||  ||||||| ||:|||||||| :|||| | |  |  ||    |: ||| 
Db    180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239

Qy    234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
          |    :||  |||:|:||:|:|||||||:||:||| ||||||  ||| ||:|||||||||
Db    240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Qy    294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352
          ||||||||||:|||||||||||||| ||||:|  ||||||| |  |||:|| : |||| |
Db    300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359

Qy    353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
          || ||||||||| ||:  | :|||||  || ||||||||:||||||:||||||| ||: |
Db    360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419

Qy    412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
          :||||||| | :|||||||||||:||:| |:||||||||||||:|||||| || |||||:
Db    420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479

Qy    472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539

Qy    532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
          ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db    540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599

Qy    592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
          |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db    600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659

Qy    652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719

Qy    712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
          |||||||| ||:||||||||||||||||| |||||||||||||||||:||||||||||||
Db    720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
```

```
Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              ||||||||||||||||||||||:|||||||:|| :|||||||||||||||||||||||
Db        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
              |:|||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
              ||||||||||||||||:||||||| || |||||||||||||||||||||||||||||||
Db        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959

Qy        952 VESSAETTDMPGQKAGT 968
              ||| ||| |  ||: ||
Db        960 VESCAETADHSGQRTGT 976
```

RESULT 3
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-2

```
  Query Match             83.9%;  Score 4332.5;  DB 9;  Length 977;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  820;  Conservative   57;  Mismatches   90;  Indels    9;  Gaps    5;

Qy          1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59
              ||||| |||||  || || || | |||:|||||||||||| ||||||||||||||    |
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116
              |||| |||:||| ||||||||:||||||| || |||   |||||||| :||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
               | :|||||| ||  |||| || :||| | |||::||||||| |:|:|| ||||| | :|
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        177 KQLSNANIGQSPHTDDHHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233
              :| ||||: ||    ||||||: || ||||||| :|||| | |: | |     |||: ||
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293
               :     : | |||||| ||: |||||| |:||:||| || |:|  ||||||:||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS-QTANKLSSCQAREADGDLRKRYLT 352
              ||||||||||:||||||||||||| | ||||:| || |||||    |  ||||  :||||
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411
              || |:||||||| ||: || :|:|| |:|||||||||:||||:|||||||||||||:  |
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471
              :||:|||||::|||||||||||||||:| |:|||||||||||||:||||| || |||||:
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531
              ||||:||||||:||||||||||:|||||||||||||||||||||||||||| |||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591
              |||||: |||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651
              ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771
              ||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831
              ||||||||||||||||||||||||||||| ||:|||||| |||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891
              |:|||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951
              |||||||||||||||||||||||| ||:|||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        952 VESSAETTDMPGQKAG 967
              |||  :||: |||: |
Db        961 VESCTQTTNQPGQRTG 976
```

RESULT 4
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-973-451-8

```
Query Match              20.8%;  Score 1074.5;  DB 9;  Length 768;
Best Local Similarity    41.7%;  Pred. No. 1.1e-76;
Matches  244;  Conservative   96;  Mismatches  194;  Indels   51;  Gaps   15;

Qy    408 RISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP--EKKWLGTPIEEMRKMPRCGIHLPS 465
Db     45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHR-NRQPFELEN 93

Qy    466 LRP-SASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTA 524
Db     94 LPPVTAGNLHRVMYQLPIRETPPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTTI 151

Qy    525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQS 584
Db    152 DFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFED 211

Qy    585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNA-KMKS 643
Db    212 LLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRKS 271

Qy    644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--LE 695
Db    272 EYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLPE 331

Qy    696 DFPEWERCEKPL--TRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLI 753
Db    332 HLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFVI 391

Qy    754 NPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEI 813
Db    392 CPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTAI 451

Qy    814 VAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR------PGVPSENLSAVATGNWGCGAF 867
Db    452 VAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGAF 503

Qy    868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRY 927
Db    504 GGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILRS 562

Qy    928 YNEECRNCSTPGP------DIKLYPFIYHAVESSAETTDMPGQKA 966
Db    563 YSRLIKEKSSKEPRENKASKKKLYDFI---KEELKKVRDVPGEGA 604
```

```
RESULT 5
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT: Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

  Query Match              11.2%;  Score 579;  DB 12;  Length 546;
  Best Local Similarity    33.9%;  Pred. No. 2.9e-37;
  Matches  150;  Conservative   69;  Mismatches  154;  Indels   70;  Gaps   12;

Qy    545 LKDAILKYNVAYS------KKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPN 598
Db     88 LADALTDLRLALALPALPPRAADGLAL--FFDDLLSRAQARGWFSEVVPNLARLLLRLPT 145

Qy    599 ICTQPIPLLKQKMNHS--------------VTMSQEQIASLLANAFFCTFPRRNAKMKSE 644
Db    146 LLEDHY----AKAGHGASGLRVMASQDAGVVLLSQELVAALLTCALFCLFPTAGRAQ--- 198

Qy    645 YSSYPDINFNRLF-----EGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSL----- 694
Db    199 -ACLPTINFDGLFAPLIHNARQSQE-QKVRCLVHYFERVTDSMPTGLVSFERKVLPRRAL 256

Qy    695 ------EDFPEWERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEE 748
Db    257 SDGVPYPDIHAWVASSAPLCQFTVFSSGFIEDEEQEALQVDFANKYLGGGALSWGCVQEE 316

Qy    749 IRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQR 808
Db    317 IRFMINPELIVGMLFLSCMEDNEAIEIFGAERFSQYMGYGSSFRFVGDYLDTKPFDSMGR 376

Qy    809 RCTEIVAIDALHFRRYLDQFVPEKVRRELNKAYCGFLR--------------------PG 848
Db    377 RRTRIVAIDALDCPARL-HYESGCLLREVNKAFCGFFDQSKHHLYAKLFQDLHNKDDFSS 435

Qy    849 VPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVV-YFTFGDSELMRDIYSMH 907
Db    436 INSSEYVGVSTGNWGCGAFGGNPEIKSMIQWIAASQALRPFVNYYTFEDVSLQR-LEEVI 494

Qy    908 TFLTERKLDVGKVYKLLLRYYNE 930
Db    495 QWIRLHGWTVGELWHMLMEYSSQ 517

RESULT 6
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-973-451-10

  Query Match               8.6%;  Score 443.5;  DB 9;  Length 726;
  Best Local Similarity    21.9%;  Pred. No. 3.1e-26;
  Matches  184;  Conservative  124;  Mismatches  278;  Indels  255;  Gaps   32;

Qy    199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258
Db     14 QDEKDYEDYV-GVGFAHQVPTM---KRRKLTEHGNTTESKEDPE--EPKSRDVFVSSQSS 67

Qy    259 DVGAEDIGTGPK----------------------NDNKLTGQESSLGD---SPPFEKES 292
```

```
Db         68 DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV 127

Qy        293 EPESPMDV-------------DNSRNSCQDSEADEETSPVF-DEQD-------------- 324
                 :| :||             :|::      :  :  :| |:|:
Db        128 NNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIEYPGIMKDTTTQ 187

Qy        325 ----DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSG 380
                  |   :|| |:   :  |||                    |||::  |   :    |
Db        188 LDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSKATKTVRTSSSS 230

Qy        381 NSSSLNVECRSSKQHGKR--DSKITDHFMRISKSE------------DRRKEQCEVRHQR 426
               |:::  |  :   |       ::  |  :  ::            ||     |  :
Db        231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLQPDLNKVDPDRNYRYCTI---- 285

Qy        427 TERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486
                          ||  |  :   |      |:  |   |:  :
Db        286 ----------PNFPASQ---GKLREDNRYGPKIVL------------------------- 307

Qy        487 PKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER-------------------TAG-- 525
                 |  :::                            |  |  |                ||
Db        308 ----PQRWREF--------------------DSRGRRRDSYFYFKRKLDGYLKCYKTTGYF 344

Qy        526 -------SRWELIQTALLNKFTRPQNLKDAILKYNVA-----YSKKWDFTALVDFWDKVL 573
                      :  ||                      |  : |  :      |  :    :    :|
Db        345 MFVGLLHNMWEF----------------DPDITYKLPALEMYYKEMSELVGREEVLEKFA 388

Qy        574 EEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT 633
                 |       :  |||:  :|    :  ::                  | |:|  :| |:|:|   ||
Db        389 RVARIAKTAEDILPE--RIYRLVGDV-------------ESATLSHKQCAALVARMFF-- 431

Qy        634 FPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TR 691
                    |:  | :|        | |:       |    |||| || ||  :::        |  |  |:|  |:
Db        432 -----ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTK 480

Qy        692 QSLEDF-PEW-ERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEI 749
                 :  |  ||  ::  :  |       |      ||        |||||     :||||      | |||||
Db        481 MDKDTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEI 539

Qy        750 RFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDGSE-K 803
              |||:  ||::|   |   |  :     |   :|   |   :|  ||||  |  :||     |  ::  :|  :
Db        540 RFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFR 599

Qy        804 DDWQRRCTEIVAIDALHFRR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSENLSAVA 858
               | :  |     | :||||: |:                  :|         : ||:  ||   ||:  |      |:  :
Db        600 DRFGRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIV 658

Qy        859 TGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVG 918
              || ||||||  ||   || :||::||   |:|  :  : :||:  ||       :     : :: :  :|
Db        659 TGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLG 718

Qy        919 K 919
              |
Db        719 K 719
```

RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(200)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

```
  Query Match              4.3%;  Score 220.5;  DB 12;  Length 200;
  Best Local Similarity   30.5%;  Pred. No. 2.9e-09;
  Matches   54;  Conservative   26;  Mismatches   68;  Indels   29;  Gaps    4;

Qy        617 MSQEQIASLLANAFFCTFPRRNAKMKSEYSSYP---------DINFNRLFEGRSSRKPEK 667
                 ::||  |   :   || ||          |  |            |::   |:|  |  :
Db         33 LTQETSAPPPSRTLFCLFP---------VSDRPGIHLPMINSDVSSGSLYEDYSQKPEST 83

Qy        668 LKTLFCYFRRVTEKKPTGLVTFTRQSLE--------DFPE---WERCEKPLTRLHVTYEG 716
                 :   |  :|:: :  | |:|:|:| |:  |            :|:       |     || |  |   |
Db         84 TGRIAHYGQRISSEMPKGIVSFERKVLPFKNDSIHISYPDANFWSTSAIPLCRFEVHSSG 143

Qy        717 TIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECL 773
               ||   |   :|| ||:::|||  |  |  ||||||:::|||      ||     :   || :
Db        144 LIEDQSSGAAEVDLANKYLGGGALGRGCVQEEIRFMVSPELEAGMLFLPAMADNEAI 200
```

RESULT 8
US-10-369-493-5013
; Sequence 5013, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5013
;   LENGTH: 6642
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-10-369-493-5013

```
  Query Match              3.4%;  Score 176;  DB 15;  Length 6642;
  Best Local Similarity   21.0%;  Pred. No. 0.0023;
  Matches  187;  Conservative  120;  Mismatches  372;  Indels  212;  Gaps   42;

Qy         21 SAPTASD-SRSFP----GRQRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRG-NATSFV- 73
              |:||   :  |  |  |      |  : :     |   :|  :      |    |     |        :   :||:  |
Db       1438 SSPTKKEKSPSSPTKKTGDEVKEKSPPKSPTKKEKSPEKPEDVKSPVKKEKSPDATNIVE 1497

Qy         74 --------FKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHK--VEKL 123
                        :  |:||   |   :    ::      |    ||       ::|       |     |   :    |     |::
Db       1498 VSSETTIEKTETTMTTEMTHESEESRTSVKKEKTPEKVDEKPKSPTKKDKSPEKSITEEI 1557

Qy        124 ENVPQLNLDKSPT--EKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPL------ 175
                               |||   |||   :  :   ::        |   ::       |        |      :|
Db       1558 ---------KSPVKKEKSPEKVEEKPASPTKKEKSPEKPASPTKKSENEVKSPTKKEKSP 1608

Qy        176 ----------PKQLSNANIGQSPHT-------------DDHSDTDHEE--DRDNQQFLT 209
                                  ||:  |           |  :                 |    |    |   |:   ::   ::   :
Db       1609 EKSVVEELKSPKEKSPEKADDKPKSPTKKEKSPEKSATEDVKSPTKKEKSPEKVEEKPTS 1668

Qy        210 PIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGP 269
               |  |   ::        |  :  :|    |      |  |:|:                    |  |  |     |
Db       1669 PTKKESSPTKKTDDEVKSPTKKEKSPQTVE--------------EKPASPTKKE------ 1708
```

```
Qy        270 KNDNKLTGQE-SSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS 328
              |:  |    :|  |   :  | : | :|:||          | :  | |:|  ||   |:    |
Db       1709 KSPEKSVVEEVKSPKEKSPEKAEEKPKSP---TKKEKSPEKSAAEEVKSPTKKEKSPEKS 1765

Qy        329 -----QTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSS 383
                   ::  |   |   :  || :::       :  |    :      |    |  :     |   :::
Db       1766 AEEKPKSPTKKESSPVKMADDEVKSPTKKEKSPEKVE-----EKPASPTKKEKTPEKSAA 1820

Qy        384 SLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC-EVRHQRTERKIPKYIPPNLPPE 442
                    | :|  :   |   |   |           |:  |   ||:   |    :: :    ||  ||   |  :
Db       1821 E---ELKSPTKKEKSPSSPT------KKTGDESKEKSPEKPEEKPKSPTPKKSPPGSPKK 1871

Qy        443 KKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKH 502
              ||    :    |   |  |        | |             : ||     ||       |     |    | |
Db       1872 KK---SKSPEAEKPP-----APKLTRDLKLQTVNKTDLAHF-EVVVEHATECKWFLDGKE 1922

Qy        503 VKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS----- 557
              :     : |  :    :|:  |              | | :     |          :::   | | |
Db       1923 I---TTAQGVTVSKDDQFEFRCS-----IDTTMFGSGT------VSVVASNAAGSVETKT 1968

Qy        558 -----------KKWDFT-----ALVDFWDKVLEEAEAQH--LYQ---------------S 584
                         || :||       |     | |   :   |  |   ||:                     :
Db       1969 ELKVLETPKETKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQNGNLLEDGKNGVT 2028

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT-----FPR--- 636
               |  :  | :| :||   |       :  : ::  |  |  |  |   |   |          |:
Db       2029 IKNEENKSSLIIPN--AQDSGKITVEASNEVG-SSESSAQLTVNPPSTTPIVVDGPKSVT 2085

Qy        637 ----RNAKMKSEYSSYP------DINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKK--PT 684
                      |: |:  | :|          ||  :: |    ||     :||   |    :::   |    |
Db       2086 IKETETAEFKATISGFPAPTVKWTIN-EKIVE--ESRTITTIKTEDVYTLKISNAKIEQT 2142

Qy        685 GLVTFTRQS-------LEDFPEWERCEKPLTRLHVTYEGTIEGNG-RGMLQVDFANRFVG 736
               | |  | |:        |           :  |   :  |   : |  :    ||     |    |::|
Db       2143 GTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLRWNLELD------- 2195

Qy        737 GGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE-CLIITGTEQYSEYTG 786
               |    |      |: :|:|  : :           :|:||   :   :|    |    |  :|
Db       2196 GPSPGT-----EVSWLLNGQPLTKSDTVQVVDHGDGTYHVTIAEAKPEMSG 2241
```

RESULT 9
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

```
  Query Match             3.4%;  Score 175;  DB 12;  Length 180;
  Best Local Similarity   28.8%;  Pred. No. 1.1e-05;
  Matches   47;  Conservative   23;  Mismatches   43;  Indels   50;  Gaps    4;

Qy        836 ELNKAYCGFL-----RP------------------------------------------ 847
              |:|||:||||      :|
Db          7 EINKAFCGFLYQCKYQPYQKILQENGCTSALFYAATSTSMETDEGEISNHKITNSQNDYH 66

Qy        848 GVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERD-VVYFTFGDSELMRDIYSM 906
              |:   |     |||||||||||||   :|  :|| |||:  |  |    : |:|||   | :: :  :
Db         67 GMDQGNNIGVATGNWGCGAFGGDPEVKTIIQWLAASQALRPFIAYYTFG-LEALQSLDEV 125

Qy        907 HTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIY 949
                 ::   ::   || :: :|: |          :    :    |     | | ||
Db        126 AHWILSQRWTVGDLWNMLIEYSINRSKGETNVGFLQWLLPSIY 168
```

RESULT 10
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233915
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_5324C.1.pep
US-10-424-599-233915

```
  Query Match             3.3%;  Score 172.5;  DB 12;  Length 300;
  Best Local Similarity   18.6%;  Pred. No. 3.8e-05;
  Matches   55;  Conservative   47;  Mismatches   64;  Indels  129;  Gaps    5;

Qy        568 FWDKVLEEAEAQHLYQSILPDMVKIALCLPNI--------------------CTQPIPLL 607
              |:|:|:     |:      :|  :||  :  |  |||::                |  : ||
Db         76 FFDEVMSGEESSKWFQEVLPALGNLLLRLPSLLESHYQNTDNMAIDGEAGAMLTTALRLL 135

Qy        608 KQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYP---------DINFNRLFE 658
                 :    |  ::||  ||:||:  :  || ||               |    |            |:  |   |::
Db        136 DSQQPGIVFLTQELIAALLSCSLFCLFP---------VSDRPVIHLPMINFDVLFGSLYD 186

Qy        659 GRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTI 718
                 | ::   |:   :    ||:|:: :  |   |:
Db        187 DYSQKQENKIWCIVHYFQRISSEMPKGIA------------------------------- 215

Qy        719 EGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGT 778
                                                                     :||  :  |  |
Db        216 ---------------------------------------------------YNERIEIVGV 225

Qy        779 EQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKV 833
              |::| || :|  ::|::|:  :|  |:|   ||  |  ::         :::  |  |:
Db        226 ERFSGYTDHASSFRFSRAKAEGREEDPVGRRKTSDLS---------EKYFPPKM 270
```

RESULT 11
US-10-115-482-18
; Sequence 18, Application US/10115482
; Publication No. US20030212257A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS
; TITLE OF INVENTION: OF USING THE SAME
; FILE REFERENCE: 21404-322D
; CURRENT APPLICATION NUMBER: US/10/115,482
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03

```
;  PRIOR APPLICATION NUMBER: 60/281,863
;  PRIOR FILING DATE: 2001-04-05
;  PRIOR APPLICATION NUMBER: 60/281,906
;  PRIOR FILING DATE: 2001-04-05
;  PRIOR APPLICATION NUMBER: 60/282,934
;  PRIOR FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: 60/283,512
;  PRIOR FILING DATE: 2001-04-12
;  PRIOR APPLICATION NUMBER: 60/285,325
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: 60/285,890
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: 60/286,068
;  PRIOR FILING DATE: 2001-04-24
;  PRIOR APPLICATION NUMBER: 60/286,292
;  PRIOR FILING DATE: 2001-04-25
;  PRIOR APPLICATION NUMBER: 60/287,213
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: 60/288,257
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: 60/291,134
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: 60/282,020
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: 60/291,725
;  PRIOR FILING DATE: 2001-05-17
;  PRIOR APPLICATION NUMBER: 60/294,771
;  PRIOR FILING DATE: 2001-05-31
;  PRIOR APPLICATION NUMBER: 60/296,965
;  PRIOR FILING DATE: 2001-06-08
;  PRIOR APPLICATION NUMBER: 60/299,128
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 149
; SEQ ID NO 18
;   LENGTH: 1489
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-115-482-18

  Query Match             3.3%;  Score 172;  DB 15;  Length 1489;
  Best Local Similarity   19.5%;  Pred. No. 0.00049;
  Matches  161;  Conservative  115;  Mismatches  301;  Indels  248;  Gaps   39;

Qy         75 KQKTITTWM---DTKGPKTAESESKENNNTRIDSMMSSVQKDNF-YPHKVEKLENVPQLN 130
              ||: :| :|        :| :  :|:|   | |  :              | :  | |  || ::
Db        418 KQRVLTNYMFPQQPRGDEDFQSDSDSFNPTLWEEQRQQRMTVAFEFEDKKEDDENAGKVK 477

Qy        131 LD-KSPTEKSSQYLNQQQ---TASVC-KWQNEGKHAEQL----------------LASEP 169
              |  ::| |:  : :  |     :    |  |     :  : :                 | |
Db        478 LSCQAPWERGQRGITLQPARLSGDCCTPWARCDQQIQDMPVPQNDPQLAWGCISGLQQER 537

Qy        170 PAGTPLPKQLSNA------------NIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTK 217
                  ||||      :              |: : |        | : ||  |    :  ::    |
Db        538 SMCTPLPVAAQSTTLPSLSGRQVEINLKRYP-------TPYPEDLKN--MVKSVQNLVGK 588

Qy        218 PTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTG 277
              |: |       ||     : : |:||:      : :  | |:     ::  ||    | |: ::
Db        589 PSHGVRVENSN-PTANTEQTVKE---KYEHKWPVAPK----EITVEDSFVHPANEMRIGE 640

Qy        278 QESSLGDSPPF----------EKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRS 327
                 || ::| :           :|||  || :|  :  |:    |     : ||    :    |
Db        641 LHPSLAETPLYPPKLVLLGKDKKESTDESEVDKTHCLNNSVSSGTYSDYSP--SQASSGS 698

Qy        328 SQTANKLSSCQA-------------------------------REA--DGDLRKR----- 349
              | |  |: | |                                |||    |:: :|
Db        699 SNTRVKVGSLQTTAKDAVHNSLWGNRIAPSFPQPLDSKPLLSQREAVPPGNIPQRPDRLP 758

Qy        350 -------YLTKGSEVRLHFQFEGENNAG--TSDLNAKPSGNSSSLNVECRSSKQHGKRDS 400
                       | ||      |:        :| |     :  |: : ::  |: :  ||:  ||:|
Db        759 MSDTFTDNWTDGS----HY-----DNTGFVAEETTAENANSNPLLSSKSRSTSSHGRRPL 809

Qy        401 KITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGT--PIEEMRKMPR 458
                 |  : :      :          | | |  |:   |    | : |   |   | |:     |
Db        810 IRQDRIVGVPLELE----------QSTHRHTPETEVPPSNPWQNWTRTPSPFEDRTAFPS 859

Qy        459 CGIHLPSLRPSASHTVTVRVDLLRAGEVPKPF----PTHYKDLWDNKHVKMPCSEQNLYP 514
                       |:  |               ::         : |:| ||      |   |  |      |: :        |: :  |
Db        860 KLETTPTTSPLPERKEHIK----ESTEIPSPFSPGVPWEYHDSNPNRSLSNVFSQIHCRP 915

Qy        515 VEDENGERTAGSRWE----LIQTALLNKFTRPQNLKDAIL----KYNVA---YSKKWDFT 563
               |   |   :      |          |::      ||| : |:: :   :           ||:       |:    |
Db        916 -ESSKGVISISKSTERLSPLMKDIKSNKFKKSQSIDEIDIGTYKVYNIPLENYASGSDHL 974

Qy        564 ALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKM--NHSVTMSQEQ 621
                   :    ||:|        :|      |:                         :|  :|:|   :|      |     |:|
Db        975 GSHERPDKML---GPEHGMSSMSR-------------SQSVPMLDDEMLTYGSSKGPQQQ 1018

Qy        622 IASLLANA--FFCTF-PRRNAKMKSE----------------------------YSSY 648
               ||:        |    :|  |: : ::|:|                             |  |
Db       1019 KASMTKKVYQFDQSFNPQGSVEVKAEKRIPPPFQHNPEYVQQASKNIAKDLISPRAYRGY 1078

Qy        649 P--------------DINFNRLFEGRSS-----RKPEKL--KTLFCYFRRVTEKKPTGL 686
              |                 |    |    |   :   :        |: :  |      |      |||| |
Db       1079 PPMEQMFSFSQPSVNEDAVVNAQFASQGARAGFLRRADSLVSATEMAMFRRVNEP----- 1133

Qy        687 VTFTRQSLEDFPEWERCEKPLTR--------LHVTYEGTIEGNGR 723
                          :  |    :|   :|  |                :  ||     ::  |||
Db       1134 --------HELPPTDRYGRPPYRGGLDRQSSVTVTESQFLKRNGR 1170

RESULT 12
US-10-282-122A-71692
; Sequence 71692, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION:  Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71692
;   LENGTH: 1297
```

```
;   TYPE: PRT
;   ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71692

  Query Match             3.3%;  Score 169.5;  DB 12;  Length 1297;
  Best Local Similarity   18.5%;  Pred. No. 0.00062;
  Matches  200;  Conservative  149;  Mismatches  411;  Indels  323;  Gaps   47;

Qy         40 DPKDAPVQF------RVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAES 93
Db        235 DSKESITKYSSSTVEHVPHESNHAIADNSNKEDTKRSSQLDSSITI------ENESTIES 288

Qy         94 ESK-ENNNTRIDSMM---SSVQKDNFYPHKV------EKLENVPQLNLDKSPTEKSSQYL 143
Db        289 SSNTSNNNERTPNYSKRDNTVNIENIYASQIVEEIRKERERKVQQKRKFKEALQNKRQQT 348

Qy        144 NQQQTASVCKWQNEGKHAEQLLA-SEPPAGTPLPKQLSNANIGQSPHT---------DDH 193
Db        349 DEEDSIQRAIDEMYAKQAEQYTGESSLNQGDNVSNKSNESEIDKSKHSYHSKDKSLADEH 408

Qy        194 -----SDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEE 248
Db        409 NRLVQNQTDEQTNSDNVDNQTEVSNESLEP------------------------YNYEE 443

Qy        249 VDVLPESPLSDVGAEDI--------GTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDV 300
Db        444 IDLNQVSSVQQVRQDDVQVKDVLEEQSSKINNNKVESYSNEKFDDYLEDTNSHEEMLHDD 503

Qy        301 DNSRNSCQDSEADEETSPVFDEQDDRSSQTAN----KLSSCQAREADGDLRKRYLTKGSE 356
Db        504 DLHEQVMDDDENEGISNKTTDENNDEKIDDANYREINESESLMQDKANDLKFNDEVNNSE 563

Qy        357 VRLHFQFEGENNAGTSDLNAKPSG-----NSSSLNVECRSSKQHGKRDSKI--------- 402
Db        564 ---NQQNSSENNINNAVRNAVSSDIEYATNEDEENDE-RLAQDTNKEDQKLSQSEDIQHE 619

Qy        403 ----------------TDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWL 446
Db        620 SLNNEDVSLTSNKTDDSEHLEKDSLNEDKKAE-------------PSFNKTNKAPQKMSI 666

Qy        447 G---------TPIEEMRKM-------PRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKP 489
Db        667 KPGSKPFNVVMTPSDKKRVMDAKKNSVSRNKVNVPELKPETKKEA-------QDEKMNAE 719

Qy        490 FPTHYKDLWDNKH-------VKMPCSEQNLYPVEDE---NGERTAGSRWELIQTALL--- 536
Db        720 FDNHLNESQLNSDESSDFNVASLEYNESSEHSVEKDNIINDENTRENEHQDVDNSQNNDM 779

Qy        537 ---NKFTRPQNLKDAILKYNVAYSKKWDFTALV--DFWDKVL------EEAEAQHLYQSI 585
Db        780 PKGNQFSKVQN-------SNNQNDNKHDINEFVSKEGYSEVTSTKNHKDGDDANH----- 827

Qy        586 LPDMVKIALCL-PNICTQPIPLLKQKMNHSVTMSQ-EQIASLLANAFFCTFPRRNAKMKS 643
Db        828 -----KAPIRRGPNIKLPSLDLLEDHEEHEIDESWIEEKKQELNDAFY------------ 870

Qy        644 EYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWE-- 701
Db        871 -YFNVP-AEVQNVTEGPSVTR----------FELAVEK---GVKVSRITALQDDIKMALA 915

Qy        702 ----RCEKP------------LTRLHVTYEGTIEG----NGRGMLQVDFANRF------ 734
Db        916 AKDIRIEAPIPGTSLVGIEVPNLNPTKVNLKSILESPKFKNAESKLTVAMGNRINNEPLL 975

Qy        735 ---------VGGGVTGAGLV---------------QEEIR-FLINPELI----------- 758
Db        976 MDIAKTPHALIAGATGSGKSVCINSILMSLLYKNHPEELRLLLIDPKMVELAPYNDLPHL 1035

Qy        759 VSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIV-AID 817
Db       1036 VSPVITDVKAATQSLKWAVDEMEKRYKLFAQFH--VRNITAFNKKAPYEQRMPKIVIVID 1093

Qy        818 AL-HFRRYLDQFVPEKVRRELNKA------------------YCGFLRPGVPSENLSAVA 858
Db       1094 ELADLMMMAPQEVEQSIARIAQKARACGIHMLVATQRPSVNVITGLIKSNIPTRIAFMVS 1153

Qy        859 TGNWGCGAFGGDARLKALIQILAAAAAER-----DVVYFTFGDSELMRDIYSMHTFLTER 913
Db       1154 S--------SVDSR-----TILDSGGAERLLGYGDMLYLGSG---MNKPIRVQGTFVSDD 1197

Qy        914 KLD 916
Db       1198 EID 1200


RESULT 13
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
;   LENGTH: 31
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-11

  Query Match             3.2%;  Score 167;  DB 9;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-06;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 792
              |||||||||||||||||||||||||||||||
Db          1 LFTEVLDHNECLIITGTEQYSEYTGYAETYR 31


RESULT 14
US-10-282-122A-71179
; Sequence 71179, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
```

```
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71179
;   LENGTH: 1169
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71179

  Query Match             3.2%;  Score 167;  DB 12;  Length 1169;
  Best Local Similarity   19.1%;  Pred. No. 0.00084;
  Matches  182;  Conservative  128;  Mismatches  290;  Indels  352;  Gaps   44;

Qy         90 TAESESKENNNTRIDSMMSSVQK-------DNFYPHKVEKLENVPQLNLDKSPTEKSSQY 142
Db        348 TGESSLDLENESNQDSSSNSLEKQSNSSNIDNKEAQNNTPLFNYEEIDLD---TTSDVYK 404

Qy        143 LNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHSDTD----H 198
Db        405 VNEEET--------ESKNDEDLVSS------------------NHYHSNDDAEVEDAEYH 438

Qy        199 EEDRDNQQFLTPIKLANTK-PTVGDGQARSNCKCSGSRQSV----------KDCTGCQQE 247
Db        439 ELDDNRQQ-----NQSNSQDDIISSKSSTSNMYDNAISASVDNNTERAKSNEDKNDTEIT 493

Qy        248 EVDVLPESPLSDVGAE---------DIGTGPKNDNKL------TGQ--ESSLGDSPPF-- 288
Db        494 HLDGTTSAKVSDEKIESNTNNHLEQDKNVKLKNVNSLKSSNSDTGQTRKQRFGGSRPFNV 553

Qy        289 ---------------EKESEPESPMDVDNSRNSCQDSEAD-----------EETSPVFD 321
Db        554 LMTPSDKKRMMDQNHKKVSVPELKPEKQANANHRKDSESNKSEEFKQINTNRETDSNSYE 613

Qy        322 E---QDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKP 378
Db        614 SNGIEHDMNSSSDEHVYETPSKQQDEQIQK--------LQDDFHFENANHAKIN--NSNE 663

Qy        379 SGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPN 438
Db        664 TGNQSNIS--------HSKRSQYSTNE----SKNIDTQTSNSSTSNQNFQR--------- 702

Qy        439 LPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDL- 497
Db        703 --------------IRKGP--NIKLPSYQ------------LL---EAPEP---HEKDQD 728

Qy        498 W-DNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAY 556
Db        729 WIDNK----------------------------------------KQELNDALYYFNV-- 746

Qy        557 SKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVT 616
Db        747 ------------------PAEVKNVTEG------------PSVTRFELSVEKGVKVSRIT 776

Qy        617 MSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFR 676
Db        777 ALQDDIKMALA-----------AK---------DIRIEAPIPGTSLVGIE---------- 806

Qy        677 RVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVG 736
Db        807 -VPNQNPTKV---NLRSIIESPKFKNTESKLT----VAMGYRINNEPLLMDIAKTPHALI 858

Qy        737 GGVTGAGLV---------------QEEIR-FLINPELI-----------VSRLFTEVLDH 769
Db        859 AGATGSGKSVCINSILMSLLYKNHPEELRLLLIDPKMVELAPYNDLPHLVSPVITDVKAA 918

Qy        770 NECLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIV-AIDAL-HFRRYLDQ 827
Db        919 TQSLKWAVEEMEKRYKLFAQYH--VRNITAFNKKAPYEQRMPKIVIVIDELADLMMMAPQ 976

Qy        828 FVPEKVRRELNKA-----------------YCGFLRPGVPSENLSAVATGNWGCGAFGG 869
Db        977 DVEQSIARIAQKARACGIHMLVATQRPSVNVITGLIKANIPTRIAFMVSS--------SV 1028

Qy        870 DARLKALIQILAAAAAER-----DVVYFTFGDSELMRDIYSMHTFLTERKLD 916
Db       1029 DSR-----TILDSGGAERLLGYGDMLYLGSG---MNKPIRVQGTFVSDDEID 1072


RESULT 15
US-10-205-823-264
; Sequence 264, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Gorbatcheva, Bella
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Wonsey, Angela M.
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Anderson, Dustin
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF PROSTATE CANCER
;  FILE REFERENCE: MRI-044
;  CURRENT APPLICATION NUMBER: US/10/205,823
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,982
;  PRIOR FILING DATE: 2001-07-25
;  PRIOR APPLICATION NUMBER: 60/314,356
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/325,020
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: 60/341,746
;  PRIOR FILING DATE: 2001-12-12
;  PRIOR APPLICATION NUMBER: 60/362,158
;  PRIOR FILING DATE: 2002-03-05
;  NUMBER OF SEQ ID NOS: 455
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 264
;   LENGTH: 1210
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-205-823-264

  Query Match             3.2%;  Score 163.5;  DB 14;  Length 1210;
  Best Local Similarity   19.7%;  Pred. No. 0.0017;
  Matches  197;  Conservative  112;  Mismatches  368;  Indels  321;  Gaps   42;

Qy         35 QRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE 94
Db        260 QDSLVAPAQPPSQTFPPPSLP-------------SKSVAMQQKPTAYVRPMDGQDQAPSE 306

Qy         95 SKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKW 154
Db        307 SPELKPLPEDYRQQTFEKTDL---KVPAKAKLTKLKMPSQSVEQT--------------Y 349
```

```
Qy    155 QNEGKHAEQLL-----------------ASEPPAGTPLP--------------------- 176

Db    350 SNEVHCVEEILKEMTHSWPPPLTAIHTPSTAEPSKFPFPTKDSQHVSSVTQNQKQYDTSS 409

Qy    177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIK--LANTKPTV-------------G 221

Db    410 KTHSNSQQGTSSMLEDDLQLSDSEDSDSEQ--TPEKPPSSSAPPSAPQSLPEPVASAHSS 467

Qy    222 DGQARSNCKCSGSRQSVKDCTGCQQEEVDVL------PESPLSDVGAEDIGTGPKNDNKL 275

Db    468 SAESESTSDSDSSSDSESESSSSDSEENEPLETPAPEPEPPTTNKW--------QLDNWL 519

Qy    276 TGQESSLGDSPPFEKESEPESPMDVDNS-----RNSCQDSEADEETSPVFDEQDDRSSQT 330

Db    520 T-------------KVSQPAAPPEGPRSTEPPRRHPESKGSSDSATSQEHSESKDPPPKS 566

Qy    331 ANKL------------SSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKP 378

Db    567 SSKAPRAPPEAPHPGKRSCQKSPAQQEPPQR-----------------QTVGTKQ-PKKP 608

Qy    379 SGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPN 438

Db    609 V--KASARAGSRTSLQ-GEREPGLLPYGSRDQTSKDKPKVKTKGRPRAAASNEPK---PA 662

Qy    439 LPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLW 498

Db    663 VPPSSE---------KKKHKSSLPAPS----------------KALSGPEPAKDNVEDRT 697

Qy    499 DNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSK 558

Db    698 PEHFALVPLTESQ-GPPHSGSGSRTSGCRQAVVVQEDSRKDRLPLPLRDT---------- 746

Qy    559 KWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQP--------------- 603

Db    747 ------------KLLSPLRDTPPPQSL---MVKITLDLLSRIPQPPGKGSRQRKAEDKQP 791

Qy    604 ----------------IPLLKQKM--------NHSVTM-----SQEQIASLLANAFFCTF 634

Db    792 PAGKKHSSEKRSSDSSSKLAKKRKGEAERDCDNKKIRLEKEIKSQSSSSSSSHKESSKTK 851

Qy    635 PRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEK--LKTLFCYFRRVTE---KKPTGLVTF 689

Db    852 PSRPSSQSSKKEMLPPPPVS-----SSSQKPAKPALKR----SRREADTCGQDPPKSASS 902

Qy    690 TRQSLED--FPEWERCEKPLTRLHVTYEGTI-------------EGNGR-GMLQVDFANR 733

Db    903 TKSNHKDSSIPKQRRVEGKGSRSSSEHKGSSGDTANPFPVPSLPNGNSKPGKPQVKFDKQ 962

Qy    734 FVGGGVTGAGL-VQEEIRFLINPELIVSRL------FTEVLDHNECLIITGTEQYSE--- 783

Db    963 -------QADLHMREAKKMKQKAELMTDRVGKAFKYLEAVLSFIECGIATESESQSSKSA 1015

Qy    784 YTGYAET------YRWARSHEDGSEKDDWQ------RRCTEIVAIDALHFRRYLDQFVPE 831

Db   1016 YSVYSETVDLIKFIMSLKSFSDATAPTQEKIFAVLCMRCQSI--LNMAMFRCKKD--IAI 1071

Qy    832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGG 869

Db   1072 KYSRTLNKHF----------ESSSKVAQAPSPCIASTG 1099
```

Search completed: May 26, 2004, 19:18:59
Job time : 83.7642 secs

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.6288 Seconds
(without alignments)
3496.721 Million cell updates/sec

Title: US-09-302-812-6
Perfect score: 5164
Sequence: 1 MSAGPGWEPCTKARWGAAGT..........YHAVESSAETTDMPGQKAGT 968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	517	10.0	997	2	B84726	probable poly(ADP-
2	443.5	8.6	726	2	T21138	hypothetical prote
3	328	6.4	364	2	A84726	probable poly(ADP-
4	176	3.4	6642	2	T29757	protein UNC-89 - C
5	170	3.3	997	2	T43523	cut17 protein - fi
6	166	3.2	1165	2	T16420	hypothetical prote
7	165	3.2	1192	2	A71623	probable secreted
8	163.5	3.2	1210	2	I39410	AF-4 protein, spli
9	163	3.2	532	2	T06029	hypothetical prote
10	161.5	3.1	2722	2	T20532	hypothetical prote
11	160	3.1	2738	2	E88320	protein F07A11.6 [
12	159.5	3.1	1115	2	T19137	hypothetical prote
13	159.5	3.1	1213	2	A58198	serine/proline-ric
14	159	3.1	1403	1	A47328	natural killer cel
15	157.5	3.0	2211	1	KFBO5	coagulation factor
16	156.5	3.0	1001	2	T16419	hypothetical prote
17	156	3.0	1274	2	T37193	enamelin matrix pr
18	156	3.0	2526	2	T20531	hypothetical prote
19	155.5	3.0	990	2	I51618	nucleolar phosphop
20	154.5	3.0	1560	2	T42727	proliferation pote
21	153	3.0	1792	2	T08878	supervillin P205 -
22	152.5	3.0	406	2	S38170	SRP40 protein - ye
23	152	2.9	489	2	A45988	dentin matrix acid
24	152	2.9	1359	2	T34036	hypothetical prote
25	150	2.9	699	2	I38073	nucleolar phosphop
26	150	2.9	1070	2	T30848	Duffy receptor - P
27	150	2.9	1337	2	T30291	dextranase - Strep
28	149.5	2.9	2361	2	T25752	hypothetical prote
29	149	2.9	1401	2	T48079	hypothetical prote
30	149	2.9	1507	2	B47328	natural killer cel
31	148.5	2.9	2706	2	T28155	variant-specific s
32	147.5	2.9	1358	2	A29360	SIR4 protein - yea
33	147.5	2.9	1611	2	T38236	hypothetical prote
34	147.5	2.9	5327	2	T13564	microtubule-associ
35	146.5	2.8	670	2	T28391	ORF MSV230 hypothe
36	146.5	2.8	1254	2	JC7185	chromosome 1 C1orf
37	146.5	2.8	2910	2	T28156	DNA-directed RNA p
38	146	2.8	1271	2	A45555	glutamate rich pro
39	146	2.8	1320	2	JC5630	TCOF1 protein - mo
40	146	2.8	1435	2	S69632	regulatory protein
41	146	2.8	1791	2	T02345	hypothetical prote
42	146	2.8	3924	2	S37431	ankyrin 2, neurona
43	145.5	2.8	559	2	B44265	ENL (translocation
44	145.5	2.8	1093	2	F88556	protein B0464.5a [
45	145.5	2.8	1463	2	T30290	AAS surface protei

ALIGNMENTS

RESULT 1
B84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-997 <STO>
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31870
A;Map position: 2

```
  Query Match             10.0%;  Score 517;  DB 2;  Length 997;
  Best Local Similarity   32.0%;  Pred. No. 7.7e-24;
  Matches  140;  Conservative   58;  Mismatches  126;  Indels  114;  Gaps   13;

Qy        569 WDKVLEEAEAQHLYQSILPDMVKIALCLPNI--------------CTQPIPLLKQKMNHS 614
              :|:::: | |::  :  |:| :  : |  |::                 : ||  :
Db         79 FDELIDEKESKRWFDEIIPALASLLLQFPSLLEVHFQNADNIVSGIKTGLRLLNSQQAGI 138

Qy        615 VTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNR--LFEGRSSRKPEKLKTLF 672
              | :||| | :||| :||| ||  |    |       | |||:     |:   |  :  |:: :
Db        139 VFLSQELIGALLACSFFCLFPDDNRGAK----HLPVINFDHLSLYISYSQSQESKIRCIM 194

Qy        673 CYFRRVTEKKPTGLVTFTRQ--SLEDFPEWERCEKPLTRLHVTYEGTIEGNGRGMLQVDF 730
               || |       | |:|:| |:  :    | :  :  |       |:            |:||
Db        195 HYFERFCSCVPIGIVSFERKITAAPDADFWSKSDVSL------YQPD------NALEVDF 242

Qy        731 ANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAET 790
              ||:::|||    | ||||||:||||||    ||     :| ||  :  |  | |::| ||||| :
Db        243 ANKYLGGGSLSRGCVQEEIRFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCYTGYASS 302

Qy        791 YRWARSHEDGSEKDDWQRRCTEIVAIDAL------HFRRYLDQFVPEKVRRELNKAYCGF 844
              :|:|  :  |    | ::|| | ||||||      ||:             : ||:||| |||
Db        303 FRFAGEYIDKKAMDPFKRRRTRIVAIDALCTPKMRHFKDIC-------LLREINKALCGF 355

Qy        845 L-----------------------RPG------------------------VPSENL- 854
              |                       ||                         |: ||
Db        356 LNCSKAWEHQNIFMDEGDNEIQLVRNGRDSGLLRTETTASHRTPLNDVEMNREKPANNLI 415

Qy        855 --------------SAVATGNWGCGAFGGDARLKALIQILAAAAAERDVV-YFTFGDSE 898
                            ||||||||| ||||  ||| || |||:    |   : |:|||  |
Db        416 RDFYVEGVDNEDHEDDGVATGNWGCGVFGGDPELKATIQWLAASQTRRPFISYYTFG-VE 474
```

```
Qy        899 LMRDIYSMHTFLTERKLD 916
               :|:: :          :||
Db        475 ALRNLDQLPPTKLSSRLD 492
```

RESULT 2
T21138
hypothetical protein F20C5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T21138
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19381
A;Accession: T21138
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-726 <WIL>
A;Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1
A;Experimental source: clone F20C5
C;Genetics:
A;Gene: CESP:F20C5.1
A;Map position: 4
A;Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

```
  Query Match             8.6%;  Score 443.5;  DB 2;  Length 726;
  Best Local Similarity   21.9%;  Pred. No. 1.7e-19;
  Matches  184;  Conservative  124;  Mismatches  278;  Indels  255;  Gaps   32;

Qy        199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258
              ::::| : ::   |:  ||:   :  |    :    : :| :|     : :   ||   | |
Db         14 QDEKDYEDYV-GVGFAHQVPTM---KRRKLTEHGNTTESKEDPE--EPKSRDVFVSSQSS 67

Qy        259 DVGAEDIGTGPK----------------------NDNKLTGQESSLGD---SPPFEKES 292
              |   ||   |:                        ||      | :|  :    :  | |::
Db         68 DESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLDNHKSTEPMEEDV 127

Qy        293 EPESPMDV-------------DNSRNSCQDSEADEETSPVF-DEQD-------------- 324
                :| :||             |::    :  :  :  :| |:|:
Db        128 NNKSNIDVAINSDEDDELVLEENNKEMRDGEQVQQLSQDLFADDQELIEYPGIMKDTTTQ 187

Qy        325 ----DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSG 380
                  |   :|| |:   :  |||                         | ||:: |    |
Db        188 LDITDSEVETAQKMEMIEETEADS-----------------TFVGEDSKATKTVRTSSSS 230

Qy        381 NSSSLNVECRSSKQHGKR--DSKITDHFMRISKSE------------DRRKEQCEVRHQR 426
               |:::  |   :   ::  |  : ::                     ||     | :
Db        231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLQPDLNKVDPDRNYRYCTI---- 285

Qy        427 TERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486
                          || |  :      |:  |   |:   :
Db        286 ----------PNFPASQ---GKLREDNRYGPKIVL------------------------- 307

Qy        487 PKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER------------------TAG-- 525
                  |  :::                           |   |                |  |
Db        308 ----PQRWREF-------------------DSRGRRRDSYFYFKRKLDGYLKCYKTTGYF 344

Qy        526 -------SRWELIQTALLNKFTRPQNLKDAILKYNVA-----YSKKWDFTALVDFWDKVL 573
                     : ||                 |  :  |  :        |  :   :    :|
Db        345 MFVGLLHNMWEF-----------------DPDITYKLPALEMYYKEMSELVGREEVLEKFA 388

Qy        574 EEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT 633
                    :  |||:  :|    : ::                 | |:|  :| |:|:|  ||
Db        389 RVARIAKTAEDILPE--RIYRLVGDV-------------ESATLSHKQCAALVARMFF-- 431

Qy        634 FPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF--TR 691
                   |:  |  :|       | |:       ||||  ||  || :::   |  | |:|  |:
Db        432 -----ARPDSPFS------FCRILSSDKSICVEKLKFLFTYFDKMSMDPPDGAVSFRLTK 480

Qy        692 QSLEDF-PEW-ERCEKPLTRLHVTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEI 749
                :  |  || ::  : |  :       |  ||          ||||||  :||||   |||||
Db        481 MDKDTFNEEWKDKKLRSLPEVEFFDEMLIEDTAL-CTQVDFANEHLGGGVLNHGSVQEEI 539

Qy        750 RFLINPELIVSRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWA-----RSHEDGSE-K 803
              |||: ||::|   |   |  :    |  :   | |  :| ||||  | :||       | :: :| :
Db        540 RFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSSYTGYGHTLKWAELQPNHSRQNTNEFR 599

Qy        804 DDWQRRCTEIVAIDALHFRR-----YLDQFVPEKVRRELNKAYCGFLRPGVPSENLSAVA 858
               | : |     | :||||: |:              :|       : ||: ||  ||:  |    |:  :
Db        600 DRFGRLRVETIAIDAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPKFTNI-PIV 658

Qy        859 TGNWGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHTFLTERKLDVG 918
              || |||||| ||   || :||::||   |:| : : :||: ||        :     : :: : :|
Db        659 TGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFCSFGEPELAAKCKKIIERMKQKDVTLG 718

Qy        919 K 919
              |
Db        719 K 719
```

RESULT 3
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

```
  Query Match             6.4%;  Score 328;  DB 2;  Length 364;
  Best Local Similarity   27.3%;  Pred. No. 8e-13;
  Matches  105;  Conservative   54;  Mismatches  123;  Indels  102;  Gaps   11;

Qy        573 LEEAEAQHLYQSILPDMVKIALCLPNICTQ--------------PIPLLKQKMNHSVTMS 618
               : : |:   :      || |   :  |   |::                   : :|       | :|
Db          1 MSKEESSRWFNEFLPAMACLLLRFPSLLESHYLNSDNLINGTKTGLRVLVPNKAGIVFLS 60

Qy        619 QEQIASLLANAFFCTFPRRNAKMKSEYSSY-PDINFNRLF-----EGRSSRKPEKLKTLF 672
              || | :||: :||| ||      :      |:: |  |||::||           ||:  :   |:| :
Db         61 QELIGALLSCSFFCLFP-----VDDRGSNHLPIINFDKLFGSLINTGRNEHQENKIKCII 115

Qy        673 CYFRRVTEKKPTGLVTFTRQ--SLE------DFPEWERCEKPLTRLHVTYEGTIEGNGRG 724
               ||:|::        | :| |:  |  |||                |   :   |    |||
Db        116 HYFQRLSSSISPGFVSFERKILSLEQDSSTLDEGFWGKSTVNLCPVEVRTSGLIEDQSVE 175

Qy        725 MLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNECLIITGTEQYSEY 784
               |:|||||: :|||    |||||||:||||||  ||      ::    |  :  : | |::| |
Db        176 ALEVDFANKNLGGGALRKGCVQEEIRFMINPELIVGMLFLPTMEVTEAIEVVGAERFSLY 235

Qy        785 TGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNKAYCGF 844
               ||     :| |:                   | ||||||
Db        236 TG---CFRKAK---------------TRIVAIDALR------------------------ 253

Qy        845 LRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYFT-------FGDS 897
                |||                           :|::|: :|  ::: | :   :          ||
Db        254 -HPGV-------------------SQYKLESLLSVLILSSSGRPIRLYMGSVSLQGIGDV 293

Qy        898 ELMRDIYSMHTFLTERKLDVGKVY 921
               || :|  |    |      :     :|
Db        294 VLMVEILSSSLFFNGLRFHRSNLY 317
```

```
RESULT 4
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

  Query Match             3.4%;  Score 176;  DB 2;  Length 6642;
  Best Local Similarity   21.0%;  Pred. No. 0.11;
  Matches  187;  Conservative  120;  Mismatches  372;  Indels  212;  Gaps   42;

Qy         21 SAPTASD-SRSFP----GRQRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRG-NATSFV- 73
Db       1438 SSPTKKEKSPSSPTKKTGDEVKEKSPPKSPTKKEKSPEKPEDVKSPVKKEKSPDATNIVE 1497

Qy         74 --------FKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHK--VEKL 123
Db       1498 VSSETTIEKTETTMTTEMTHESEESRTSVKKEKTPEKVDEKPKSPTKKDKSPEKSITEEI 1557

Qy        124 ENVPQLNLDKSPT--EKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPL------ 175
Db       1558 ---------KSPVKKEKSPEKVEEKPASPTKKEKSPEKPASPTKKSENEVKSPTKKEKSP 1608

Qy        176 ----------PKQLSNANIGQSPHT-------------DDHSDTDHEE--DRDNQQFLT 209
Db       1609 EKSVVEELKSPKEKSPEKADDKPKSPTKKEKSPEKSATEDVKSPTKKEKSPEKVEEKPTS 1668

Qy        210 PIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGP 269
Db       1669 PTKKESSPTKKTDDEVKSPTKKEKSPQTVE--------------EKPASPTKKE------ 1708

Qy        270 KNDNKLTGQE-SSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS 328
Db       1709 KSPEKSVVEEVKSPKEKSPEKAEEKPKSP---TKKEKSPEKSAAEEVKSPTKKEKSPEKS 1765

Qy        329 -----QTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSS 383
Db       1766 AEEKPKSPTKKESSPVKMADDEVKSPTKKEKSPEKVE-----EKPASPTKKEKTPEKSAA 1820

Qy        384 SLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC-EVRHQRTERKIPKYIPPNLPPE 442
Db       1821 E---ELKSPTKKEKSPSSPT------KKTGDESKEKSPEKPEEKPKSPTPKKSPPGSPKK 1871

Qy        443 KKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKH 502
Db       1872 KK---SKSPEAEKPP-----APKLTRDLKLQTVNKTDLAHF-EVVVEHATECKWFLDGKE 1922

Qy        503 VKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS----- 557
Db       1923 I---TTAQGVTVSKDDQFEFRCS-----IDTTMFGSGT------VSVVASNAAGSVETKT 1968

Qy        558 -----------KKWDFT-----ALVDFWDKVLEEAEAQH--LYQ---------------S 584
Db       1969 ELKVLETPKETKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQNGNLLEDGKNGVT 2028

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT-----FPR--- 636
Db       2029 IKNEENKSSLIIPN--AQDSGKITVEASNEVG-SSESSAQLTVNPPSTTPIVVDGPKSVT 2085

Qy        637 ----RNAKMKSEYSSYP------DINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKK--PT 684
Db       2086 IKETETAEFKATISGFPAPTVKWTIN-EKIVE--ESRTITTIKTEDVYTLKISNAKIEQT 2142

Qy        685 GLVTFTRQS-------LEDFPEWERCEKPLTRLHVTYEGTIEGNG-RGMLQVDFANRFVG 736
Db       2143 GTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLRWNLELD------- 2195

Qy        737 GGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE-CLIITGTEQYSEYTG 786
Db       2196 GPSPGT-----EVSWLLNGQPLTKSDTVQVVDHGDGTYHVTIAEAKPEMSG 2241

RESULT 5
T43523
cut17 protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Jun-2000
C;Accession: T43523; T41649; T41700
R;Morishita, J.; Matsusaka, T.; Yanagida, M.
submitted to the EMBL Data Library, August 1999
A;Description: Fission yeast cut17 is required for chromosome segregation.
A;Reference number: Z22536
A;Accession: T43523
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-997 <MOR>
A;Cross-references: EMBL:AB031034; PIDN:BAA83415.1
R;Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22007
A;Accession: T41649
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-997 <HAR>
A;Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC962.02c
A;Experimental source: strain 972h-; cosmid c962
R;Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z22010
A;Accession: T41700
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 932-997 <WED>
A;Cross-references: EMBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10c
C;Genetics:
A;Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c
A;Map position: 3L
A;Introns: 43/3

  Query Match             3.3%;  Score 170;  DB 2;  Length 997;
  Best Local Similarity   21.2%;  Pred. No. 0.017;
  Matches  135;  Conservative   82;  Mismatches  227;  Indels  194;  Gaps   32;

Qy          7 WEP-------------CTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPP 52
Db        173 WEPDDDPYTEHKRRRADCVFFTWKDPNSLSPT---KLSFLSTSN--IDPEDLTEDNSILP 227

Qy         53 SSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAE-SESKENN-NTRIDSMM--- 107
Db        228 VSP-----------------------TRDSTKSHKTLNFSPSRKNNLNARPLTMSLYT 262

Qy        108 -SSVQKDNFYPHKVEKLENVPQL-----NLDKSPTEKSSQYLNQQQTASVCK-----WQN 156
Db        263 NTSEEKDS-QPTRAPQSPTKPVLLTAPRRKNKSP---------KKSKPAVFKPVKPIFSD 312

Qy        157 EGKHAEQLLASEPPAGTPLPKQLSN--ANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLA 214
Db        313 EDEDDDDLTASQ-----PFSKGICNDSMQVAKKNFTEEIPLKEDEKDNELEHLVSPATSV 367

Qy        215 NTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNK 274
```

```
Db    368 HT-------------------TVSDITGHQS---------VTDESDE---------- 386
Qy    275 LTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSSQTANKL 334
Db    387 ---QNNCMSTPPKIEIESKIEEEISVVS-----KSKEISSSVSSVGKEQNHTEKQVAIET 438
Qy    335 SSCQAREADGDLRKRYLTKGSEVRLHFQFEGE--NNAGTSDLNAKPSGNSSSLNVECRSS 392
Db    439 PEQQKVEKEDE-------------HLNLQGSFIEESTKQPISSKPSTSSPDMTDAATGG 484
Qy    393 KQHGK--RDS---------KITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP 441
Db    485 RVSSSSFRDKILQTNFSPRSTIDSFSNISKK--RNSEEANDENDETNLKIP-------IP 535
Qy    442 EKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNK 501
Db    536 EKKRKFQEVLQSKNI------LVSSTEDSHEPVKVTEDSQTAIHVSK-----FEDL-ENK 583
Qy    502 HVKMPCSEQNLYPV-EDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS--- 557
Db    584 SME---SEQSLQLLSESENDDK------PLIDLIPLLAIKRKDNLVSGVLEKGKSTSTSK 634
Qy    558 KKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALC 595
Db    635 TKFD-TSIVDFIEKPKTE------ISEVLPEEKRKAIC 665
```

RESULT 6
T16420
hypothetical protein F52C9.8b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16420
R;Favello, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F52C9.
A;Reference number: Z18511
A;Accession: T16420
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1165 <FAV>
A;Cross-references: EMBL:U39850; NID:g1055052; PID:g1055056; PIDN:AAA81058.1; CESP:F52C9
C;Genetics:
A;Gene: CESP:F52C9.8b
A;Introns: 21/1; 36/1; 61/3; 79/3; 107/3; 464/3; 566/1; 1010/3; 1029/3; 1106/3

```
  Query Match             3.2%;  Score 166;  DB 2;  Length 1165;
  Best Local Similarity   19.7%;  Pred. No. 0.038;
  Matches  146;  Conservative   94;  Mismatches  251;  Indels  250;  Gaps   33;

Qy     32 PGRQRRVLDPKDAPVQFRVPPSSPACV---SGRAGPHRGNATSFVFKQKTITTWMDTKGP 88
Db    478 PPQQQRQKQQRSQPAPARVPPQVPSQVPVTGGVAADEPPPPCSY-------------SP 523
Qy     89 KTAESESK-------------------------ENNNTRIDSMMSSVQKDNFYPHKVE 121
Db    524 VAQSSESKIEPVDVKPRVAPVPPQVPVTPTKPVITNNKKKRID----VVTLDEDAPRRVQ 579
Qy    122 KLENVPQLN-------LDKSPTEKSSQYLNQQQTASVC---------KWQ---------- 155
Db    580 VKQEIPEVSSTSDATKSDAAPTARGAVRIKQEVESDVAPNTILISAKKFERMKAEAEDKE 639
Qy    156 ----------------NEGKHAEQLLASEPPAGTPLPKQ--LSNANIGQ----------- 186
Db    640 DMKKKIAALQEALFNIQEERRVEKEIAAFATTNQAVPQNQPASSVQIAQVSTSESDAPGT 699
Qy    187 ----------SPHTDDH----SDTDHEEDRDNQQFLTPIKLANTK----------PTVG 221
Db    700 SEAAATETMTSPKTKNNVIVETEGEQEEDEDE----IPIKKSKKRRAKIVSNDEEEEPVR 755
Qy    222 DGQARSNCKCS----GSRQSVKDCTGCQQEEVDVLPESP------LSDVGAEDIGTGPKN 271
Db    756 HPKRRSDEKREKRHVSYAESDDDMPVVKKKRRNQSPEDPEYSAASPSEDEDDDIGSFVVS 815
Qy    272 DNKLTGQES-SLGDSPPFEKESEPESPM-DVDNSRNSCQDSEADEETSPVFDEQDDRSSQ 329
Db    816 DNEDDDADSFVVGDDEPIEYEEEDEDDMIERRSSRKRRSDSRSKKSATPT-DRRRSRDTP 874
Qy    330 TANKLSSCQAREADGDLRKRYLTKGS-EVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVE 388
Db    875 TGSR--SMRSTSPNDRRKSRETPPGNRSMRRTSPSDGRKSRDTP--TASSSMSSSTLSY- 929
Qy    389 CRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERK----------------- 430
Db    930 CKKSKETPMSYEEIE------QQKKAKRQRNCKTREENRERKRLAQLEELESSETTGVRR 983
Qy    431 IPKYIPPNLPPEKKWLGTPIEEMRKMPR----------------------------C 459
Db    984 TLRSTQDNSDPLDASLATTIEEFRKTKKKDAKSSENRAKEKQKPMNKRPTSSASVDSNDD 1043
Qy    460 GIHLPSLRPSASHTVTVRVDLLRAGEVPKPF------------PTHYKDLWDNKHVKMP 506
Db   1044 GVHIPAKR-----------MAHASSVPGPSRSKPPMIGAVKNRPNH-TEMLDKRNKE-- 1088
Qy    507 CSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNL-----KDAILKYNV------A 555
Db   1089 -SEEKRRKDRDEL-ERLRNKKHTTEEEKI--KMARLQNALKVVGKAAGLKATVKKELTGS 1144
Qy    556 YSKKWDFTALV-DFWDKVLEE 575
Db   1145 PAKKQKPAPLCRKFWISRLEE 1165
```

RESULT 7
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71623
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Shallom, S.; Mason, T.; Yu, K.; Fujii, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.; Lai, Z.; Schwartz, D.C.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1192 <GAR>
A;Cross-references: GB:AE001373; GB:AE001362; NID:g3845097; PIDN:AAC71813.1; PID:g38450
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0115w

```
  Query Match             3.2%;  Score 165;  DB 2;  Length 1192;
  Best Local Similarity   21.5%;  Pred. No. 0.045;
  Matches   85;  Conservative   61;  Mismatches  143;  Indels  106;  Gaps   18;

Qy     70 TSFVFKQKTITTWMD-TKGPKTAESESKENNNTRIDSMMSSVQKDNF-YPHKVEKLEN-V 126
Db    235 SDFMNKMKYQNTQIDNNKGKKTTNTMEKNKKNRDKKHSKKRKTKQNYKYKKENQNIENHI 294
Qy    127 PQLN--------LDKSPTE-KSSQYLNQQQ-------TASVC-----KWQNEGKHAEQLL 165
Db    295 PQSKYKQERIEILDDNGKELKSHKNIKEEKGGIEKTDTTNIADIKIKKEERETKDEKEKN 354
Qy    166 ASEPPAGTPLPKQLSNANIGQSPH---------TDDHSDTDHEEDRDNQQFLTPIKLANT 216
Db    355 IQQ------LVKDVQLIKVGEETKDDEKEDKEGTDDEEDTDDEEDTDDEEDTDDEEDTSD 408
Qy    217 KPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLT 276
Db    409 EETTGDQE--------------------NKEETEV--------------------DEKKT 428
Qy    277 --GQESSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSSQTANKL 334
Db    429 EKAEEELEEDKEESEKDKE-ESEKDKEESEKDKEESEKDKEKT----EEDEEKTED-EKG 482
```

```
Qy        335 SSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQ 394
                  :| | | :|                    :||   || |              | :  ::
Db        483 TEVYKKETDVDEKKE--------------KGEYGEGTDDEEDKEKEEDDE---ETKVEEK 525

Qy        395 HGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTER 429
                ::| : ||:       |:  : |: :|  ::|||
Db        526 KTEKDEEGTDYEEDTDDSD--KDEETKVEEKKTER 558
```

RESULT 8
I39410
AF-4 protein, splice form 3 - human
N;Alternate names: serine/proline-rich FEL protein
C;Species: Homo sapiens (man)
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997
C;Accession: I39410
R;Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Gale, R.P.; Lang
Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993
A;Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leuk
A;Reference number: A47440; MUID:93281633; PMID:8506309
A;Accession: I39410
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1210 <NAK>
A;Cross-references: GB:L13773; NID:g306446; PID:g306447
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit
C;Genetics:
A;Gene: GDB:MLLT2; AF-4; FEL
A;Cross-references: GDB:136792; OMIM:159557
A;Map position: 4q21-4q21
C;Keywords: alternative splicing

```
  Query Match             3.2%;  Score 163.5;  DB 2;  Length 1210;
  Best Local Similarity   19.7%;  Pred. No. 0.057;
  Matches  197;  Conservative  112;  Mismatches  368;  Indels  321;  Gaps   42;

Qy         35 QRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE 94
              | :: |   | |  |||  |                 : |   :||         |  ||
Db        260 QDSLVAPAQPPSQTFPPPSLP------------SKSVAMQQKPTAYVRPMDGQDQAPSE 306

Qy         95 SKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKW 154
              | |      |    : :| :     ||      : :| :     |::              :
Db        307 SPELKPLPEDYRQQTFEKTDL---KVPAKAKLTKLKMPSQSVEQT-------------Y 349

Qy        155 QNEGKHAEQLL----------------ASEPPAGTPLP-------------------- 176
                ||    |::|                       ::   |:   | |
Db        350 SNEVHCVEEILKEMTHSWPPPLTAIHTPSTAEPSKFPFPTKDSQHVSSVTQNQKQYDTSS 409

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIK--LANTKPTV------------G 221
              | ||:  |  |    :|        || ||::|  ||  |     ::   |:
Db        410 KTHSNSQQGTSSMLEDDLQLSDSEDSDSEQ--TPEKPPSSSAPPSAPQSLPEPVASAHSS 467

Qy        222 DGQARSNCKCSGSRQSVKDCTGCQQEEVDVL------PESPLSDVGAEDIGTGPKNDNKL 275
               :: |       |  |  :  :      ||  : |            || | ::           : || |
Db        468 SAESESTSDSDSSSDSESESSSSDSEENEPLETPAPEPEPPTTNKW--------QLDNWL 519

Qy        276 TGQESSLGDSPPFEKESEPESPMDVDNS-----RNSCQDSEADEETSPVFDEQDDRSSQT 330
              |               | |:| :| :  |     |:         :|  ||    |  |   ::
Db        520 T-------------KVSQPAAPPEGPRSTEPPRRHPESKGSSDSATSQEHSESKDPPPKS 566

Qy        331 ANKL-------------SSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKP 378
              ::|               |||    |      :|                    ||     ||
Db        567 SSKAPRAPPEAPHPGKRSCQKSPAQQEPPQR-----------------QTVGTKQ-PKKP 608

Qy        379 SGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPN 438
                 :|        |:|  |:|: |:|:  :  :  : |  :            ||    |
Db        609 V--KASARAGSRTSLQ-GEREPGLLPYGSRDQTSKDKPKVKTKGRPRAAASNEPK---PA 662

Qy        439 LPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLW 498
              :||   :          :|   :   :   ||                :|     |:|     : :|
Db        663 VPPSSE---------KKKHKSSLPAPS----------------KALSGPEPAKDNVEDRT 697

Qy        499 DNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSK 558
                   :| :|      |        :| ||:|  |   ::          |      |  |:|
Db        698 PEHFALVPLTESQ-GPPHSGSGSRTSGCRQAVVVQEDSRKDRLPLPLRDT---------- 746

Qy        559 KWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQP--------------- 603
                                |:|            ||:   |||| | |  :  ||
Db        747 ------------KLLSPLRDTPPPQSL---MVKITLDLLSRIPQPPGKGSRQRKAEDKQP 791

Qy        604 ----------------IPLLKQKM--------NHSVTM-----SQEQIASLLANAFFCTF 634
                                | |::          |    : :     ||    :|        |
Db        792 PAGKKHSSEKRSSDSSSKLAKKRKGEAERDCDNKKIRLEKEIKSQSSSSSSHKESSKTK 851

Qy        635 PRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEK--LKTLFCYFRRVTE---KKPTGLVTF 689
              | |  :    |:              :     ||:||  |   ||       ||   :     :  |        :
Db        852 PSRPSSQSSKKEMLPPPPVS-----SSSQKPAKPALKR----SRREADTCGQDPPKSASS 902

Qy        690 TRQSLED--FPEWERCEKPLTRLHVTYEGTI------------EGNGR-GMLQVDFANR 733
              |: : :|     |:  |  |   :|    ::|:                    || :  |   || |   :
Db        903 TKSNHKDSSIPKQRRVEGKGSRSSSEHKGSSGDTANPFPVPSLPNGNSKPGKPQVKFDKQ 962

Qy        734 FVGGGVTGAGL-VQEEIRFLINPELIVSRL------FTEVLDHNECLIITGTEQYSE--- 783
                          | | ::|   :        ||:    |:                 ||    || |  |  :|  |
Db        963 -------QADLHMREAKKMKQKAELMTDRVGKAFKYLEAVLSFIECGIATESESQSSKSA 1015

Qy        784 YTGYAET------YRWARSHEDGSEKDDWQ------RRCTEIVAIDALHFRRYLDQFVPE 831
              |: |:||              :|   |   :            |  |   :  :    ||     |    :
Db       1016 YSVYSETVDLIKFIMSLKSFSDATAPTQEKIFAVLCMRCQSI--LNMAMFRCKKD--IAI 1071

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGG 869
               |   | |||  :                   |:  |  ||    |  |
Db       1072 KYSRTLNKHF----------ESSSKVAQAPSPCIASTG 1099
```

RESULT 9
T06029
hypothetical protein T28I19.100 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06029
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z15484
A;Accession: T06029
A;Molecule type: DNA
A;Residues: 1-532 <BEV>
A;Cross-references: EMBL:AL035709; GSPDB:GN00062; ATSP:T28I19.100
A;Experimental source: cultivar Columbia; BAC clone T28I19
C;Genetics:
A;Gene: ATSP:T28I19.100
A;Map position: 4

```
  Query Match             3.2%;  Score 163;  DB 2;  Length 532;
  Best Local Similarity   23.1%;  Pred. No. 0.019;
  Matches   83;  Conservative   61;  Mismatches  145;  Indels   70;  Gaps   17;

Qy         92 ESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASV 151
              | ||| |:   :        || :|   | | | ||           :  : ||    |  ||:|
Db        189 EGESKSNSTENV-----SVHEDESGP-KNEVLEGSVIKEVSLNTTENGSDDGEQQETKS- 241

Qy        152 CKWQNEGKHAEQLLASEPPAGTPLPK-QLSNANI------------GQSPHTDDHSDTDH 198
                : :  |  |:  :     :       |:   || :|            | |   :    :  |  :
Db        242 ---ELDSKTGEKGFSD---SNGELPETNLSTSNATETTESSGSDESGSSGKSTGYQQTKN 295

Qy        199 EEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLS 258
              |||           |     |   |  :   |||   :   |: |   |    :   ||
Db        296 EEDE--------------KEKVQSSEEESKVKESGKNE--KDASSSQDESKEEKPERKKK 339

Qy        259 DVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSP 318
               :  :    | | : : :   :| |              ::||:  | |   : :   :| |:     :||
Db        340 EESSSQ-GEGKEEEPEKREKEDSSS-----QEESKEEEPENKEKEASSSQEENEIKETE- 392

Qy        319 VFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGT------- 371
                |::: |||   |:      :  : ::      :|          || ::    |  |   ::: |
```

```
Db        393 -IKEKEESSSQEGNENKETEKKSSES---QRKENTNSEKKIE-QVESTDSSNTQKGDEQK 447

Qy        372 SDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQ------CEVRH 424

Db        448 TDESKRESGNDTS-NKE--TEDDSSKTESEKKEENNRNGETEETQNEQEQTKSALEISH 503
```

RESULT 10
T20532
hypothetical protein F07A11.6b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T20532; T27777
R;Palmer, S.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19287
A;Accession: T20532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2722 <WIL>
A;Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b
A;Experimental source: clone F07A11
R;Gajadsty, S.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z20417
A;Accession: T27777
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2722 <WI2>
A;Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b
A;Experimental source: clone ZK20
C;Genetics:
A;Gene: CESP:F07A11.6b
A;Map position: 2
A;Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1

```
  Query Match            3.1%;  Score 161.5;  DB 2;  Length 2722;
  Best Local Similarity  19.0%;  Pred. No. 0.24;
  Matches  150;  Conservative  116;  Mismatches  265;  Indels  259;  Gaps   35;

Qy          9 PCTKARWGAAGTSAPTAS--DSRSFPGRQRRVLDPKDAPVQFRVPPS------------ 53

Db       1416 PRQRNRTTSSTSTATTSSKHEALSIPEK------PLSPPVTAKSSVSSIDDPSIRDEFSM 1469

Qy         54 -----SPACVSGR-------------AGPHRGNATSFVFKQKTITTWMDTKGPKTAESE 94

Db       1470 NSAADSPMSTTGRPMVLTKAAMKAFNSTPPKKKNSSS--------------GQHDSSSG 1514

Qy         95 SKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEK--------SSQYLNQQ 146

Db       1515 SSSDSSSSDGSTSSDDSSDDEVPKQTEPVTSIPVVASDNGSPENVVVETPSIVSQTPREP 1574

Qy        147 QTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIG--QSPHTDDHSDTDHEEDRDN 204

Db       1575 EPFTISEQSSE---SEPEAVPECPEASVEPQMETSQNVEPVSEEHEDSHEHGDSEVAVES 1631

Qy        205 QQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAED 264

Db       1632 QQ--QPLE--------------------------------HQEEKEEL-ENKILDVAAE- 1655

Qy        265 IGTGPKNDNKLTGQESSLGDS-PPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQ 323

Db       1656 -----HHEEQVQGDEDSVESSIPAPSDEPDPVTQAEKSAHTLISDQETDQAVQSIFDEE 1710

Qy        324 D-DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNS 382

Db       1711 EADEFPQYPDFGISTNEKEVSG--KDPHNIKPTE---------PLNNGHTDLLFSPSSSA 1759

Qy        383 SSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPP--NLP 440

Db       1760 HA------SEKQS--------------TKSEDDMEEDSELVVMEKEVPMEQVIAQEVHVP 1799

Qy        441 PEKKWLGTPIEEMRKMPRCGIHLPSLRP---SASHTVTVRVDLLRAGE---VPKPFPTHY 494

Db       1800 SEP----SPMEEEVKLETSPV--PKEEPIKMEESPEQTPTPDLISNNESQDTPGAVNNHL 1853

Qy        495 KDLWD----------NKH-VKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRP 542

Db       1854 HENHDAVQTPIQLQPASQHQVAQPSPRPAVAPDSQQNG-----------PVLVSQQSQP 1901

Qy        543 QNL--KDAILKYNVAYSKKWDFTALVDFWDKVLEEAEA------------QHLY----- 582

Db       1902 SPMSSQQSDMAQNLILSSK----DINDLAAKLHKNPEALAQATRGDCSGIFQHLLLHAQG 1957

Qy        583 --QSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAK 640

Db       1958 NGQNMTPEMLQLKAAF---------FAQQQENEANQMMQ------------------AK 1989

Qy        641 MKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEW 700

Db       1990 MKQQ-----TINKDRI------KEQERVKRMYEENERKVE-------------EDRREK 2024

Qy        701 ERCEKPLTRL 710

Db       2025 QRKEEERQRL 2034
```

RESULT 11
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: E88320
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_el
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
A;Accession: E88320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2738 <STO>
A;Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A1
C;Genetics:
A;Gene: F07A11.6
A;Map position: 2

```
  Query Match            3.1%;  Score 160;  DB 2;  Length 2738;
  Best Local Similarity  19.4%;  Pred. No. 0.3;
  Matches  132;  Conservative  104;  Mismatches  235;  Indels  210;  Gaps   31;

Qy         84 DTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEK----- 138

Db       1526 DSSSGSSSDSSSSDGSTSSDDS------SDDEVPKQTEPVTSIPVVASDNGSPENVVVET 1579

Qy        139 ---SSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIG--QSPHTDDH 193

Db       1580 PSIVSQTPREPEPFTISEQSSE---SEPEAVPECPEASVEPQMETSQNVEPVSEEHEDSH 1636

Qy        194 SDTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLP 253

Db       1637 EHGDSEVAVESQQ--QPLE-----------------------------HQEEKEEL- 1661

Qy        254 ESPLSDVGAEDIGTGPKNDNKLTGQESSLGDS-PPFEKESEPESPMDVDNSRNSCQDSEA 312

Db       1662 ENKILDVAAE------HHEEQVQGDEDSVESSIPAPSDEPDPVTQAEKSAHTLISDQET 1715

Qy        313 DEETSPVFDEQD-DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGT 371

Db       1716 DQAVQSIFDEEEADEFPQYPDFGISTNEKEVSG--KDPHNIKPTE---------PLNNGH 1764

Qy        372 SDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKI 431

Db       1765 TDLLFSPSSSAHA------SEKQS--------------TKSEDDMEEDSELVVMEKEVPM 1804
```

```
Qy        432 PKYIPP--NLPPEKKWLGTPIEEMRKMPRCGIHLPSLRP---SASHTVTVRVDLLRAGE- 485
              :  |      ::|  |      :|:||  |:     :    |          ||:   |
Db       1805 EQVIAQEVHVPSEP----SPMEEEVKLETSPV--PKEEPIKMEESPEQTPTPDLISNNES 1858

Qy        486 --VPKPFPTHYKDLWD-----------NKH-VKMPCSEQNLYPVEDENGERTAGSRWELI 531
                |       :    |            ::|  |    |      :  |    :||
Db       1859 QDTPGAVNNHLHENHDAVQTPIQLQPASQHQVAQPSPRPAVAPDSQQNG----------- 1907

Qy        532 QTALLNKFTRPQNL--KDAILKYNVAYSKKWDFTALVDFWDKVLEEAEA----------- 578
                 |::: ::|   :   :  :  :   |:   ||       :  |     |:  :   ||
Db       1908 -PVLVSQQSQPSPMSSQQSDMAQNLILSSK----DINDLAAKLHKNPEALAQATRGDCSG 1962

Qy        579 --QHLY-------QSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANA 629
                |||          |::  |:|:::                  :|:  |   :   |  |
Db       1963 IFQHLLLHAQGNGQNMTPEMLQLKAAF---------FAQQQENEANQMMQ---------- 2003

Qy        630 FFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTF 689
                           |||| :        || :|:         :: |::|  ::       |     |
Db       2004 ---------AKMKQQ-----TINKDRI------KEQERVKRMYEENERKVE--------- 2034

Qy        690 TRQSLEDFPEWERCEKPLTRL 710
                      ||  |  :|  |:   ||
Db       2035 -----EDRREKQRKEEERQRL 2050
```

RESULT 12
T19137
hypothetical protein C09G1.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19137
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19079
A;Accession: T19137
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1115 <WIL>
A;Cross-references: EMBL:Z50176; PIDN:CAA90539.1; GSPDB:GN00028; CESP:C09G1.2
A;Experimental source: clone C09G1
C;Genetics:
A;Gene: CESP:C09G1.2
A;Map position: X
A;Introns: 14/1; 49/2; 1095/2

```
  Query Match             3.1%;  Score 159.5;  DB 2;  Length 1115;
  Best Local Similarity   21.5%;  Pred. No. 0.09;
  Matches  167;  Conservative  109;  Mismatches  297;  Indels  203;  Gaps   40;

Qy         76 QKTITT-WMDTKGPKTAESES----------------KENNNTRIDSMM----------- 107
              || :::   :|:|   |:|:||:                 ::   | | |   :
Db        289 QKQVSSEEVDSKNQKSAQSETDLNYFTLNNNSSKTLFEQQRNVRTDQDLNNNRDSCSACG 348

Qy        108 --SSVQKDNFY-PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCK----------- 153
                |  :|::       |   |||         :   |  |||||    :   |  :    |
Db        349 NNSDAEKEDIKGTRKRSSLENQHSERMQFSQTEKSSPIGDMVQNSPDFKNMANRTPAESE 408

Qy        154 --------WQN-----EGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHSDTDHEE 200
                      ||:         |: |                  : ||  :   |       |  |
Db        409 KGKKKAQDWQHSLNFRSGRSA-------------INKQQNEIKIKPS------SSNDQKR 449

Qy        201 D-----RDNQQFLTPIKLANTKPTVGDGQARSNCKC-SGSRQSVKDCTGCQ-QEEVDVLP 253
                      |::  |       :|     ||   |        |    | |  | | : ||     :: | | |
Db        450 HWLFKMRESAQTKKNLKYLIAKRTHQSMQCQSEFKCVSSSNQTEKDLKSFSIKKNVQVQP 509

Qy        254 ESP---LSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQDS 310
                ||     ||:|  |           |::   |   |||:    |||                :  :   :
Db        510 CSPSLKSSDIGNE--------RNRIKRQHFSLGEQ-SFEKRKP---------ALTATGNE 551

Qy        311 EADEETSPVFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAG 370
               :  | :        :      : ||    |   |    ||:: :::       |  :::    |:    |
Db        552 KGDLKRKLSLLKDQFSAKQTEQSL-QCSADLEDGEIDEKHCQILSNSQVNRVFKKRKTA- 609

Qy        371 TSDLNAKPSGNSSSLNVECRS-SKQHGKRDSKITDHFMRISKSE------DRRKEQCEVR 423
               | ||    |          | | | |    |     |      ::|    |          :|||     :
Db        610 ---LAAKEPNN---FTVSCCSDSNSTGITDDPIQTKNTKLSLEEQDPLLSERRKTLATIF 663

Qy        424 HQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPS----------LRPSA--- 470
               :|       |  |  |  ||             |:|        :|  |  |          | |:|
Db        664 NQ----KKPVQIDQNLQCSFDLEDGEIDE----KQCQISLHSQVHRVLKKRQLAPAAKDL 715

Qy        471 -SHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVED--ENGERTAGSR 527
                : ||    :|             :|      |    :|     :       |          |     |      |
Db        716 NNFTVNCSLDSNSNESNDEPIKTEKTQLLLKEHQPVISKRQETLKTESSFNNFNSTDAHR 775

Qy        528 WELIQTALLNKFTRPQNLKDAILKYN--------VAY------SKKWDFTA-------LV 566
                   |:  |:  |    :|  |:|:|:: :  |                :  |        ::  |  :  :           ||
Db        776 NCRRQSKLIEK--QPISLEDSVFEQNKSAQPESVINYFTIDNKNQDWQYFSGPEKDPKLV 833

Qy        567 DFWDKV-LEEAEAQHLYQSILPDMVKIALCLP---------NICTQPIPLLKQKMNHSVT 616
                | :  |   ::|   :|    :       |          |           |           :  |::      ||      |
Db        834 SFLEAVNKKDALVKHYSEKQRPVRNKNDSDYPTATDTDQRSSSCSE----LK-----GTT 884

Qy        617 MSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPD---INFNRLFEGRSSRKPEKLKTLFC 673
              |:::|    :                          |:     :|:|  :|  |:     :        :    : :|  |:   :|
Db        885 MNKKQEEN----------TRKRLSLKNEEAS-PEKTLVRAEKDFNHFTVKKKEECLSLSF 933

Qy        674 YFRRVTEKKPTGLVTFTRQSLE---DFPEWERCEKPLTRLHVT-YEGT-IEGNGRG 724
                |   :   ||                        |:|           :  ||  :      :  :|:  |:   |   :: ||||
Db        934 GFIALPIKKRRPFFEGQNNSIELKGTATQIER-QSVSSEVHISNYQSTPMKNNGRG 988
```

RESULT 13
A58198
serine/proline-rich FEL protein, splice form 1 - human
N;Alternate names: AF-4 protein
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 25-Oct-1996 #text_change 01-Dec-2000
C;Accession: A58198; I52572
R;Morrissey, J.J.M.
Blood 81, 1124-1131, 1993
A;Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.
A;Reference number: I52572; MUID:93184301; PMID:8443374
A;Accession: A58198
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1213 <MOR>
A;Cross-references: GB:L25050; NID:g407323
A;Accession: I52572
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'MAAQS',13-1213 <MO2>
A;Cross-references: GB:L25050; NID:g407323; PIDN:AAA36642.1; PID:g407324
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit]
C;Genetics:
A;Gene: GDB:MLLT2; AF-4; FEL
A;Cross-references: GDB:136792; OMIM:159557
A;Map position: 4q21-4q21
C;Keywords: alternative splicing

```
  Query Match             3.1%;  Score 159.5;  DB 2;  Length 1213;
  Best Local Similarity   19.7%;  Pred. No. 0.1;
  Matches  196;  Conservative  109;  Mismatches  367;  Indels  321;  Gaps   42;

Qy         35 QRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE 94
              |     :: |    |     |    |||  |                 :  |     :||           |     | ||
Db        267 QDSLVAPAQPPSQTFPPPSLP-------------SKSVAMQQKPTAYVRPMDGQDQAPSE 313

Qy         95 SKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKW 154
              |  |        |       :  :|  :      ||         :  :|   :         |::                          :
Db        314 SPELKPLPEDYRQQTFEKTDL---KVPAKAKLTKLKMPSQSVEQT--------------Y 356

Qy        155 QNEGKHAEQLL-----------------ASEPPAGTPLP--------------------- 176
               ||       |::|                           ::    |:        |   |
Db        357 SNEVHCVEEILKEMTHSWPPPLTAIHTPSTAEPSKFPFPTKDSQHVSSVTQNQKQYDTSS 416
```

```
Qy   177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIK--LANTKPTV-------------G 221
Db   417 KTHSNSQQGTSSMLEDDLQLSDSEDSDSEQ--TPEKPPSSSAPPSAPQSLPEPVASAHSS 474

Qy   222 DGQARSNCKCSGSRQSVKDCTGCQQEEVDVL------PESPLSDVGAEDIGTGPKNDNKL 275
Db   475 SAESESTSDSDSSSDSESESSSSDSEENEPLETPAPEPEPPTTNKW--------QLDNWL 526

Qy   276 TGQESSLGDSPPFEKESEPESPMDVDNS-----RNSCQDSEADEETSPVFDEQDDRSSQT 330
Db   527 T-------------KVSQPAAPPEGPRSTEPPRRHPESKGSSDSATSQEHSESKDPPPKS 573

Qy   331 ANKL------------SSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKP 378
Db   574 SSKAPRAPPEAPHPGKRSCQKSPAQQEPPQR-----------------QTVGTKQ-PKKP 615

Qy   379 SGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPN 438
Db   616 V--KASARAGSRTSLQ-GGREPGLLPYGSRDQTSKDKPKVKTKGRPRAAASNEPK---PA 669

Qy   439 LPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLW 498
Db   670 VPPSSE---------KKKHKSSLPAPS----------------KALSGPEPAKDNVEDRT 704

Qy   499 DNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSK 558
Db   705 PEHFALVPLTESQ-GPPHSGSGSRTSGCRQAVVVQEDSRKDRLPLPLRDT---------- 753

Qy   559 KWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQP--------------- 603
Db   754 ------------KLLSPLRDTPPPQSL---MVKITLDLLSRIPQPPGKGSRQRKAEDKQP 798

Qy   604 ----------------IPLLKQKM--------NHSVTM-----SQEQIASLLANAFFCTF 634
Db   799 PAGKKHSSEKRSSDSSSKLAKKRKGEAERDCDNKKIRLEKEIKSQSSSSSSHKESSKTK 858

Qy   635 PRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEK--LKTLFCYFRRVTE---KKPTGLVTF 689
Db   859 PSRPSSQSSKKEMLPPPPVS-----SSSQKPAKPALKR----SRREADTCGQDPPKVPAV 909

Qy   690 TRQSLED--FPEWERCEKPLTRLHVTYEGTI-------------EGNGR-GMLQVDFANR 733
Db   910 PRVNHKDSSIPKQRRVEGKGSRSSSADKGSSGDTANPFPVPSLPNGNSKPGKPQVKFDKQ 969

Qy   734 FVGGGVTGAGL-VQEEIRFLINPELIVSRL------FTEVLDHNECLIITGTEQYSE--- 783
Db   970 -------QADLHMREAKKMKQKAELMTDRVGKAFKYLEAVLSFNECGIATESESQSSKSA 1022

Qy   784 YTGYAET------YRWARSHEDGSEKDDWQ------RRCTEIVAIDALHFRRYLDQFVPE 831
Db  1023 YSVYSETVDLIKFIMSLKSFSDATAPTQEKIFAVLCMRCQSI--LNMAMFRCKKD--IAI 1078

Qy   832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGC 864
Db  1079 KYSRTLNKHF----------ESSSKVAQAPSPC 1101
```

RESULT 14
A47328
natural killer cell tumor-recognition protein - human
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumc
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C;Accession: A47328
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824; PMID:8421688
A;Accession: A47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1403 <AND>
A;Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252
A;Experimental source: NK killer cells from adult blood
A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)
C;Genetics:
A;Gene: GDB:NKTR
A;Cross-references: GDB:137171; OMIM:161565
A;Map position: 3p23-3p21
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>

```
  Query Match          3.1%;  Score 159;  DB 1;  Length 1403;
  Best Local Similarity   19.2%;  Pred. No. 0.13;
  Matches  146;  Conservative   99;  Mismatches  252;  Indels  262;  Gaps   32;

Qy    20 TSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTI 79
Db   591 TASKSSSHSRS----------------------RSKSRSSSKSG-HRKRA---------- 617

Qy    80 TTWMDTKGPKTAESESKENNNTRIDSMMSSV-QKDNFYPHKVEKLENVPQLNLDKSP--- 135
Db   618 -----SKSPRKTASQLSENKPVKTEPLRATMAQNENVVVQPV-VAENIPVIPLSDSPPPS 671

Qy   136 ------------------TEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
Db   672 RWKPGEKPWKPSYERIQEMKAKTTHLLPIQSTYSLANIKETG-----------SSSSYH 719

Qy   177 KQLSNANIGQSPHT--DDHSDTDHEEDRD--------NQQFLTPIKLANTKPTVGDGQAR 226
Db   720 KREKNSESDQSTYSKYSDRSSESSPRSRSRSSRSRSYSRSYTRSRSLASS-----HSRSR 774

Qy   227 SNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGD-- 284
Db   775 SPSSRSHSRNKYSDHSQCSR-------SSSYTSISSDD-GRRAKRRLRSSGKKNSVSHKK 826

Qy   285 ------------------SPPFEKESEPESPMDV--DNSRNSCQDSEADEETSPVFDEQ 323
Db   827 HSSSSEKTLHSKYVKGRDRSSCVRKYSESRSSLDYSSDSEQSSVQATQSAQE-----KEK 881

Qy   324 DDRSSQTANKLS-------------------------SCQAREADGDLRKRY------- 350
Db   882 QGQMERTHNKQEKNRGEEKSKSERECPHSKKRTLKENLSDHLRNGSKPKRKNYAGSKWDS 941

Qy   351 -------LTKGSEVRLHFQFEGENNAGTSD---------LNAKPSGNSS---SL---NVE 388
Db   942 ESNSERDVTKNSKNDSHPSSDKEEGEATSDSESEVSEIHIKVKPTTKSSTNTSLPDDNGA 1001

Qy   389 CRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRH---QRTERKIPKYIPPNLPPEKKW 445
Db  1002 WKSSKQR----TSTSDSEGSCSNSENNRGKPQKHKHGSKENLKREHTKKVKEKLKGKK-- 1055

Qy   446 LGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKM 505
Db  1056 -----DKKHKAPK--------RKQAFH-----------WQPPLEFGEEEEEEIDDKQVTQ 1091

Qy   506 PCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVA----YSKKWD 561
Db  1092 ESKEKKV----SENNE---------------------TIKDNILKTEKSSEEDLSGKHD 1125

Qy   562 FTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQ 621
Db  1126 TVTVSSDLDQFTKDDSKLSISPTALNTEENVA-CLQNI---------QHVEESVPNGVED 1175

Qy   622 IASLLANAFFCTFPRRNAKMKSEYSS--------YPDIN 652
Db  1176 VLQTDDNMEICT-PDRSSPAKVEETSPLGNARLDTPDIN 1213
```

RESULT 15
KFBO5
coagulation factor V precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C;Accession: A42580; A36497

```
R;Guinto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A;Title: The complete cDNA sequence of bovine coagulation factor V.
A;Reference number: A42580; MUID:92147638; PMID:1737753
A;Accession: A42580
A;Molecule type: mRNA
A;Residues: 1-2211 <GUI>
A;Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038
A;Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBIP:80776)
R;Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A;Title: Identification and characterization of a phospholipid-binding site of bovine fa
A;Reference number: A36497; MUID:91072354; PMID:2254316
A;Accession: A36497
A;Molecule type: protein
A;Residues: 1566-1570,'X',1572-1581,'X',1583-1584;1673-1676,'X',1678-1679,'X',1681,'X',1
R;Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A;Title: Determination of the disulfide bridges in factor Va heavy chain.
A;Reference number: A55979; MUID:95034740; PMID:7947716
A;Contents: annotation
A;Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
C;Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C;Function:
A;Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A;Pathway: blood coagulation
C;Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C;Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-2211/Product: coagulation factor V #status predicted <MAT>
F;29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F;29-345/Domain: A1 <DA1>
F;33-329/Domain: ferroxidase repeat homology <FO1>
F;346-695/Domain: A2 <DA2>
F;351-688/Domain: ferroxidase repeat homology <FO2>
F;696-1564/Domain: B <DOB>
F;1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F;1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F;1565-1892/Domain: A3 <DA3>
F;1572-1892/Domain: ferroxidase repeat homology <FO3>
F;1654-1752/Region: phospholipid binding #status predicted
F;1893-2051/Domain: C1 <DC1>
F;1893-2048/Domain: discoidin I amino-terminal homology <DN1>
F;2052-2211/Domain: C2 <DC2>
F;2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F;167-193,248-329,499-525/Disulfide bonds: #status experimental
F;225,239,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,1094,
F;334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F;363,697,1537/Binding site: sulfate (Tyr) (covalent) #status predicted
F;376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F;533-534/Cleavage site: Arg-Gly (protein C) #status predicted
F;607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
F;741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F;1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F;1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

  Query Match             3.0%;  Score 157.5;  DB 1;  Length 2211;
  Best Local Similarity   18.7%;  Pred. No. 0.31;
  Matches  135;  Conservative  102;  Mismatches  295;  Indels  189;  Gaps   28;

Qy         65 HRGNATSFVFKQKTITTWMDTKGP----------------------------------- 88

Db        641 HEDTLTLFPMQGESVTVTMDNVGTWMLTTMNSNPRSKKLRLRFRDAKCIRNDDDDSYEII 700

Qy         89 -----KTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKL------------ENVPQLNL 131

Db        701 YEPSGSTAMTTKKIHDSSEIEDENDA---DSDYQDELALILGLRSFRNSSLNQEKDELNL 757

Qy        132 DKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASE-------------PPAGTPLPKQ 178

Db        758 TALALEKDSEFIPPSANRSLDSNSSSRSHVSRLIAKNFAESLKTLLHLEAPAAGSPLEHA 817

Qy        179 LSNANIGQSPHTDDHSDTDHEEDRDNQQF--LTPIKL-------ANTKPT------VGDG 223

Db        818 GLDKNSALNPPMAEHSSPYSEDPREDHPLSDVTGVSLLPFGTGFKNRKPAKHQRFQVGRG 877

Qy        224 QARSNCKCSGSR-QSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNK-LTGQESS 281

Db        878 QAAKH-KFSQTRFPAHKTRTRLSQDNSSSSRMGPWEDIPSDLLLLQQKDPYKILNGEWHL 936

Qy        282 LGDSPPFEKESEPESPMDVDNSRNSCQ-------------DSEADEETSPVF-------- 320

Db        937 VSEKGSYEIIQDANENKTVNKLPNSPQNDSRTWGENIPFKNSHGKQSGHPTFLVTRRKPL 996

Qy        321 -DEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPS 379

Db        997 QDRQDRRNSRLKEGLPLIRTRRKKKEEKPAYHVPLSPRSFH-PLRGEVNASFSDRRHNHS 1055

Qy        380 -----GNSSSLNVECRSS--KQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIP 432

Db       1056 LLLHASNETSLSIDLNQTFPSMNLSLAASLPDH------------DQTSPNDTTSQTSSP 1103

Qy        433 KYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLR-PSASHTVTV---RVDLLRAGEVPK 488

Db       1104 PDLYPTVSPEEHYQIFPIQD--SDPTHSTTAPSNRSPDPTHSTTAPSNRSPPTQPSQIPN 1161

Qy        489 PFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDA 548

Db       1162 ---------YDLRNRAIPTDVSQIFP----------SLELEVWQTATSLDLSQP------ 1196

Qy        549 ILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPI-PLL 607

Db       1197 --------------SISPDLGQMALSPDPGQ---ESLSPDLGQTSLS-PDLSQESLSPDL 1238

Qy        608 KQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLF--EGRSSRKP 665

Db       1239 GQTA-LSPDPSQESLSPDLGQTALSPDP-------SQESLSPDLGQTALSPDPGQESLSP 1290

Qy        666 E 666

Db       1291 D 1291

Search completed: May 26, 2004, 18:47:56
Job time : 50.6288 secs
```

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:          May 26, 2004, 17:20:49 ; Search time 15.5091 Seconds
                                          (without alignments)
                                          3249.964 Million cell updates/sec

Title:           US-09-302-812-6
Perfect score:   5164
Sequence:        1 MSAGPGWEPCTKARWGAAGT..........YHAVESSAETTDMPGQKAGT 968

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	3.4	6632	1	UN89_CAEEL	O01761 caenorhabdi
2	171.5	3.3	1647	1	PQE1_CAEEL	Q10124 caenorhabdi
3	170	3.3	997	1	BIR1_SCHPO	O14064 schizosacch
4	163.5	3.2	1210	1	AF4_HUMAN	P51825 homo sapien
5	162	3.1	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
6	161	3.1	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
7	159	3.1	687	1	DSPP_RAT	Q62598 rattus norv
8	159	3.1	1462	1	NKCR_HUMAN	P30414 homo sapien
9	157.5	3.0	2211	1	FA5_BOVIN	Q28107 bos taurus
10	156	3.0	1274	1	ENAM_MOUSE	O55196 mus musculu
11	154	3.0	934	1	DSPP_MOUSE	P97399 mus musculu
12	153	3.0	5085	1	PCLO_RAT	Q9jks6 rattus norv
13	152.5	3.0	406	1	SR40_YEAST	P32583 saccharomyc
14	152.5	3.0	5120	1	PCLO_CHICK	Q9pu36 gallus gall
15	152	2.9	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
16	150	2.9	489	1	DMP1_RAT	P98193 rattus norv
17	150	2.9	699	1	NP14_HUMAN	Q14978 homo sapien
18	150	2.9	1070	1	PVDR_PLAVS	P22290 plasmodium
19	150	2.9	1337	1	DEXT_STRDO	P39653 streptococc
20	149.5	2.9	513	1	DMP1_HUMAN	Q13316 homo sapien
21	149.5	2.9	2314	1	AKA6_RAT	Q9wvc7 rattus norv
22	149	2.9	1453	1	NKCR_MOUSE	P30415 mus musculu
23	147.5	2.9	1358	1	SIR4_YEAST	P11978 saccharomyc
24	147.5	2.9	5596	1	MDN1_HUMAN	Q9nu22 homo sapien
25	146	2.8	1395	1	SP41_YEAST	P38904 saccharomyc
26	146	2.8	3924	1	ANK2_HUMAN	Q01484 homo sapien
27	145.5	2.8	559	1	ENL_HUMAN	Q03111 homo sapien
28	145.5	2.8	1093	1	YKD5_CAEEL	Q03563 caenorhabdi
29	145.5	2.8	1282	1	BMS1_HUMAN	Q14692 homo sapien
30	145.5	2.8	1447	1	BUD4_YEAST	P47136 saccharomyc
31	145	2.8	574	1	MIG1_CANAL	Q9y7g2 candida alb
32	143.5	2.8	503	1	DMP1_MOUSE	O55188 mus musculu
33	143	2.8	1253	1	DSPP_HUMAN	Q9nzw4 homo sapien
34	141.5	2.7	1489	1	YGP0_YEAST	P53115 saccharomyc
35	141	2.7	1031	1	YDG9_SCHPO	Q10496 schizosacch
36	140.5	2.7	646	1	SG1_BOVIN	P23389 bos taurus
37	140.5	2.7	1087	1	POM1_SCHPO	Q09690 schizosacch
38	140.5	2.7	3259	1	GOB1_HUMAN	Q14789 homo sapien
39	140	2.7	623	1	HPC2_YEAST	Q01448 saccharomyc
40	139.5	2.7	1088	1	MRTB_HUMAN	Q9ulh7 homo sapien
41	139.5	2.7	1176	1	KMLS_BOVIN	Q28824 bos taurus
42	139.5	2.7	2738	1	PGCV_RAT	Q9erb4 rattus norv
43	139.5	2.7	5560	1	SPEN_DROME	Q8sx83 drosophila
44	139	2.7	925	1	PIP1_YEAST	P40020 saccharomyc
45	139	2.7	950	1	STL2_MOUSE	Q99n50 mus musculu

ALIGNMENTS

```
RESULT 1
UN89_CAEEL
ID   UN89_CAEEL     STANDARD;      PRT;  6632 AA.
AC   O01761; Q17362;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN   UNC-89 OR C09D1.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=Bristol N2;
RX   MEDLINE=96180278; PubMed=8603916;
RA   Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT   "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT   assembly, encodes a giant modular protein composed of Ig and signal
RT   transduction domains.";
RL   J. Cell Biol. 132:835-848(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z., Le T.T., Wilson R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC       lattice assembly begins with positional cues laid down in the
CC       basement membrane and muscle cell membrane. UNC-89 responds to
CC       these signals, localizes, and then participates in assembling an
CC       M-line.
CC   -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC   -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 5 RCSD domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33058; AAB00542.1; -.
DR   EMBL; AF003131; AAB54132.2; -.
DR   PDB; 1FHO; 20-DEC-00.
DR   WormPep; C09D1.1; CE30426.
```

```
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR007850; RCSD.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 47.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF05177; RCSD; 5.
DR   Pfam; PF00621; RhoGEF; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 23.
DR   SMART; SM00325; RhoGEF; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW   3D-structure.
FT   DOMAIN        63    127       SH3.
FT   DOMAIN       152    330       DH.
FT   DOMAIN       342    498       PH.
FT   DOMAIN       547    633       IG-LIKE C2-TYPE 1.
FT   DOMAIN       648    736       IG-LIKE C2-TYPE 2.
FT   DOMAIN       748    838       IG-LIKE C2-TYPE 3.
FT   DOMAIN       946   1033       IG-LIKE C2-TYPE 4.
FT   DOMAIN      1044   1132       IG-LIKE C2-TYPE 5.
FT   DOMAIN      1140   1227       IG-LIKE C2-TYPE 6.
FT   DOMAIN      1272   1315       THR-RICH.
FT   DOMAIN      1375   1475       RCSD 1.
FT   DOMAIN      1479   1585       RCSD 2.
FT   DOMAIN      1597   1695       RCSD 3.
FT   DOMAIN      1700   1799       RCSD 4.
FT   DOMAIN      1800   1860       RCSD 5.
FT   DOMAIN      1982   2067       IG-LIKE C2-TYPE 7.
FT   DOMAIN      2071   2163       IG-LIKE C2-TYPE 8.
FT   DOMAIN      2171   2261       IG-LIKE C2-TYPE 9.
FT   DOMAIN      2269   2359       IG-LIKE C2-TYPE 10.
FT   DOMAIN      2367   2455       IG-LIKE C2-TYPE 11.
FT   DOMAIN      2463   2564       IG-LIKE C2-TYPE 12.
FT   DOMAIN      2563   2651       IG-LIKE C2-TYPE 13.
FT   DOMAIN      2657   2746       IG-LIKE C2-TYPE 14.
FT   DOMAIN      2754   2858       IG-LIKE C2-TYPE 15.
FT   DOMAIN      2887   2980       IG-LIKE C2-TYPE 16.
FT   DOMAIN      2994   3081       IG-LIKE C2-TYPE 17.
FT   DOMAIN      3087   3183       IG-LIKE C2-TYPE 18.
FT   DOMAIN      3189   3280       IG-LIKE C2-TYPE 19.
FT   DOMAIN      3286   3376       IG-LIKE C2-TYPE 20.
FT   DOMAIN      3384   3469       IG-LIKE C2-TYPE 21.
FT   DOMAIN      3482   3572       IG-LIKE C2-TYPE 22.
FT   DOMAIN      3580   3667       IG-LIKE C2-TYPE 23.
FT   DOMAIN      3686   3777       IG-LIKE C2-TYPE 24.
FT   DOMAIN      3817   3908       IG-LIKE C2-TYPE 25.
FT   DOMAIN      3920   4009       IG-LIKE C2-TYPE 26.
FT   DOMAIN      4018   4106       IG-LIKE C2-TYPE 27.
FT   DOMAIN      4109   4201       IG-LIKE C2-TYPE 28.
FT   DOMAIN      4212   4297       IG-LIKE C2-TYPE 29.
FT   DOMAIN      4302   4387       IG-LIKE C2-TYPE 30.
FT   DOMAIN      4400   4485       IG-LIKE C2-TYPE 31.
FT   DOMAIN      4489   4580       IG-LIKE C2-TYPE 32.
FT   DOMAIN      4588   4678       IG-LIKE C2-TYPE 33.
FT   DOMAIN      4681   4771       IG-LIKE C2-TYPE 34.
FT   DOMAIN      4873   4961       IG-LIKE C2-TYPE 35.
FT   DOMAIN      4965   5057       IG-LIKE C2-TYPE 36.
FT   DOMAIN      5067   5160       IG-LIKE C2-TYPE 37.
FT   DOMAIN      5171   5260       IG-LIKE C2-TYPE 38.
FT   DOMAIN      5277   5366       IG-LIKE C2-TYPE 39.
FT   DOMAIN      5383   5472       IG-LIKE C2-TYPE 40.
FT   DOMAIN      5487   5578       IG-LIKE C2-TYPE 41.
FT   DOMAIN      5595   5685       IG-LIKE C2-TYPE 42.
FT   DOMAIN      5701   5790       IG-LIKE C2-TYPE 43.
FT   DOMAIN      5815   5904       IG-LIKE C2-TYPE 44.
FT   DOMAIN      5925   6014       IG-LIKE C2-TYPE 45.
FT   DOMAIN      6038   6130       IG-LIKE C2-TYPE 46.
FT   DOMAIN      6150   6239       IG-LIKE C2-TYPE 47.
FT   DOMAIN      6275   6368       FIBRONECTIN TYPE-III.
FT   DOMAIN      6413   6502       IG-LIKE C2-TYPE 48.
FT   DOMAIN      6507   6596       IG-LIKE C2-TYPE 49.
FT   DISULFID     568    621       POTENTIAL.
FT   DISULFID    2908   2975       POTENTIAL.
FT   DISULFID    3015   3065       POTENTIAL.
FT   DISULFID    3707   3759       POTENTIAL.
FT   DISULFID    3826   3890       POTENTIAL.
FT   DISULFID    5092   5157       POTENTIAL.
FT   DISULFID    5298   5350       POTENTIAL.
FT   DISULFID    5508   5560       POTENTIAL.
FT   DISULFID    5616   5669       POTENTIAL.
FT   DISULFID    5722   5764       POTENTIAL.
FT   DISULFID    5836   5901       POTENTIAL.
FT   DISULFID    5946   5998       POTENTIAL.
FT   DISULFID    6036   6171       POTENTIAL.
FT   DISULFID    6421   6486       POTENTIAL.
FT   CONFLICT    2137   2137       A -> P (IN REF. 1).
FT   CONFLICT    2245   2247       AKA -> PKP (IN REF. 1).
FT   CONFLICT    2258   2258       A -> P (IN REF. 1).
FT   CONFLICT    2284   2284       E -> G (IN REF. 1).
FT   CONFLICT    2297   2297       M -> I (IN REF. 1).
FT   CONFLICT    3531   3531       A -> G (IN REF. 1).
FT   CONFLICT    3884   3888       DAGEY -> RRRRI (IN REF. 1).
FT   CONFLICT    3929   3929       A -> V (IN REF. 1).
FT   CONFLICT    5134   5134       A -> P (IN REF. 1).
FT   CONFLICT    5145   5145       T -> S (IN REF. 1).
FT   CONFLICT    5185   5185       G -> A (IN REF. 1).
FT   CONFLICT    5199   5199       K -> N (IN REF. 1).
FT   CONFLICT    5202   5202       L -> F (IN REF. 1).
FT   CONFLICT    5213   5213       F -> L (IN REF. 1).
FT   CONFLICT    6178   6178       A -> G (IN REF. 1).
FT   CONFLICT    6268   6268       K -> E (IN REF. 1).
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;

  Query Match             3.4%;  Score 176;  DB 1;  Length 6632;
  Best Local Similarity   21.0%;  Pred. No. 0.12;
  Matches  187;  Conservative  120;  Mismatches  372;  Indels  212;  Gaps   42;

Qy         21 SAPTASD-SRSFP----GRQRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRG-NATSFV- 73

Db       1438 SSPTKKEKSPSSPTKKTGDEVKEKSPPKSPTKKEKSPEKPEDVKSPVKKEKSPDATNIVE 1497

Qy         74 --------FKQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHK--VEKL 123

Db       1498 VSSETTIEKTETTMTTEMTHESEESRTSVKKEKTPEKVDEKPKSPTKKDKSPEKSITEEI 1557

Qy        124 ENVPQLNLDKSPT--EKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPL------ 175

Db       1558 ---------KSPVKKEKSPEKVEEKPASPTKKEKSPEKPASPTKKSENEVKSPTKKEKSP 1608

Qy        176 ----------PKQLSNANIGQSPHT-------------DDHSDTDHEE--DRDNQQFLT 209

Db       1609 EKSVVEELKSPKEKSPEKADDKPKSPTKKEKSPEKSATEDVKSPTKKEKSPEKVEEKPTS 1668

Qy        210 PIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGP 269

Db       1669 PTKKESSPTKKTDDEVKSPTKKEKSPQTVE--------------EKPASPTKKE------ 1708

Qy        270 KNDNKLTGQE-SSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSS 328

Db       1709 KSPEKSVVEEVKSPKEKSPEKAEEKPKSP---TKKEKSPEKSAAEEVKSPTKKEKSPEKS 1765

Qy        329 -----QTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSS 383

Db       1766 AEEKPKSPTKKESSPVKMADDEVKSPTKKEKSPEKVE-----EKPASPTKKEKTPEKSAA 1820
```

```
Qy        384 SLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQC-EVRHQRTERKIPKYIPPNLPPE 442
                  |:|  :   |  |  |           |:  |  ||:  |    :: :     ||  ||  | :
Db       1821 E---ELKSPTKKEKSPSSPT------KKTGDESKEKSPEKPEEKPKSPTPKKSPPGSPKK 1871

Qy        443 KKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKH 502
              ||    :    |  | |      | |                 : ||     ||      |     |     | |
Db       1872 KK---SKSPEAEKPP-----APKLTRDLKLQTVNKTDLAHF-EVVVEHATECKWFLDGKE 1922

Qy        503 VKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS----- 557
              :     :  |  :    :|:   |           | | :     |          ::: |  | |
Db       1923 I---TTAQGVTVSKDDQFEFRCS-----IDTTMFGSGT------VSVVASNAAGSVETKT 1968

Qy        558 -----------KKWDFT-----ALVDFWDKVLEEAEAQH--LYQ---------------S 584
                          || :||          |       |   |   :   |   |   ||:                    :
Db       1969 ELKVLETPKETKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQNGNLLEDGKNGVT 2028

Qy        585 ILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCT-----FPR--- 636
              |   :   | :| :||    |      :   : ::  |   |   |   |   |    |          |:
Db       2029 IKNEENKSSLIIPN--AQDSGKITVEASNEVG-SSESSAQLTVNPPSTTPIVVDGPKSVT 2085

Qy        637 ----RNAKMKSEYSSYP------DINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKK--PT 684
                      |: |:   | :|          ||   :    ||      :||   |   :::   |     |
Db       2086 IKETETAEFKATISGFPAPTVKWTIN-EKIVE--ESRTITTIKTEDVYTLKISNAKIEQT 2142

Qy        685 GLVTFTRQS-------LEDFPEWERCEKPLTRLHVTYEGTIEGNG-RGMLQVDFANRFVG 736
               | |  |  |:           |           :  |   :  :| :   ||    |   |::|
Db       2143 GTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLRWNLELD------- 2195

Qy        737 GGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE-CLIITGTEQYSEYTG 786
               |    |          |: :|:|  : :      :|:||  :     :|   |    | :|
Db       2196 GPSPGT-----EVSWLLNGQPLTKSDTVQVVDHGDGTYHVTIAEAKPEMSG 2241


RESULT 2
PQE1_CAEEL
ID   PQE1_CAEEL     STANDARD;      PRT;  1647 AA.
AC   Q10124; Q8MQ26; Q8MQ27; Q8MQ28; Q8MQ29; Q8MQ30; Q8MQ31;
DT   01-FEB-1996 (Rel. 33, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   PolyQ enhancer protein 1.
GN   PQE-1 OR F52C9.8/F52C9.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND C), FUNCTION, AND SUBCELLULAR
RP   LOCATION.
RX   MEDLINE=22388320; PubMed=12486229;
RA   Faber P.W., Voisine C., King D.C., Bates E.A., Hart A.C.;
RT   "Glutamine/proline-rich PQE-1 proteins protect Caenorhabditis elegans
RT   neurons from huntingtin polyglutamine neurotoxicity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17131-17136(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Favello A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Waterston R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Protects neurons from the toxic effects of expanded
CC       poly-Q disease proteins. It is not known if this is done via
CC       participation in the pathogenic mechanism underlying poly-Q-
CC       induced neurodegeneration or acting as a genetic modifier of the
CC       age of onset or progression of neurodegeneration.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC       Name=b; Synonyms=Pqe-1A;
CC         IsoId=Q10124-1; Sequence=Displayed;
CC       Name=a; Synonyms=Pqe-1C;
CC         IsoId=Q10124-2; Sequence=VSP_008069, VSP_008070;
CC       Name=c; Synonyms=Pqe-1B;
CC         IsoId=Q10124-3; Sequence=VSP_008068;
CC       Name=d;
CC         IsoId=Q10124-4; Sequence=VSP_008065;
CC         Note=No experimental confirmation available;
CC       Name=e;
CC         IsoId=Q10124-5; Sequence=VSP_008063, VSP_008064;
CC         Note=No experimental confirmation available;
CC       Name=f;
CC         IsoId=Q10124-6; Sequence=VSP_008066, VSP_008067;
CC         Note=No experimental confirmation available;
CC   -!- DOMAIN: The Gln/Pro-rich N-terminus and the Arg/Asp/Glu/Lys-rich
CC       charged domain are critical in protecting glutamatergic ASH
CC       sensory neurons from degeneration. ASH neurons expressing isoforms
CC       lacking these domains show progressive degeneration.
CC   -!- SIMILARITY: To mammalian protein GOR, yeast YGR276C and X.laevis
CC       XPMC2.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U39850; AAM45366.1; -.
DR   EMBL; U39850; AAM45367.2; -.
DR   EMBL; U39850; AAM45368.1; -.
DR   EMBL; U39850; AAM45369.1; -.
DR   EMBL; U39850; AAM45370.1; -.
DR   EMBL; U39850; AAM45371.1; -.
DR   PIR; T16421; T16421.
DR   WormPep; F52C9.8a; CE30802.
DR   WormPep; F52C9.8b; CE30998.
DR   WormPep; F52C9.8c; CE30672.
DR   WormPep; F52C9.8d; CE30673.
DR   WormPep; F52C9.8e; CE30674.
DR   WormPep; F52C9.8f; CE30675.
DR   InterPro; IPR006055; Exonuclease.
DR   Pfam; PF00929; Exonuclease; 1.
DR   SMART; SM00479; EXOIII; 1.
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN       33    568       GLN/PRO-RICH.
FT   DOMAIN      614   1356       ARG/ASP/GLU/LYS-RICH.
FT   DOMAIN     1478   1636       EXONUCLEASE.
FT   VARSPLIC      1   1074       Missing (in isoform e).
FT                                /FTId=VSP_008063.
FT   VARSPLIC   1075   1091       DAKSSENRAKEKQKPMN -> MNVSSPSTQASHFNQFQ
FT                                (in isoform e).
FT                                /FTId=VSP_008064.
FT   VARSPLIC      1   1089       Missing (in isoform d).
FT                                /FTId=VSP_008065.
FT   VARSPLIC      1   1207       Missing (in isoform f).
FT                                /FTId=VSP_008066.
FT   VARSPLIC   1208   1299       AKKQKPAPAVPKILDFSVGRTFTAIRQTAIKLVFDTFLERD
FT                                SPNAAREAQEFELSIAKQYTDGQKYRINIGHKVAALRKENT
FT                                SGILEVNKNA -> MSTYYQSAPVLVHAHAYQPPLSPPQSP
FT                                PHGYLPFYTSNPMMCSTSPLYFSPQMVQPLPPQQHYHMYPS
FT                                TSAGPGMYYTDFNPHYVQIQISQS (in isoform f).
FT                                /FTId=VSP_008067.
FT   VARSPLIC     96   1072       Missing (in isoform c).
FT                                /FTId=VSP_008068.
FT   VARSPLIC   1073   1081       KKDAKSSEN -> VRRCGGDNV (in isoform a).
FT                                /FTId=VSP_008069.
FT   VARSPLIC   1082   1647       Missing (in isoform a).
FT                                /FTId=VSP_008070.
SQ   SEQUENCE   1647 AA;  182261 MW;  A3A66BA183B3D650 CRC64;

  Query Match              3.3%;  Score 171.5;  DB 1;  Length 1647;
```

```
Best Local Similarity   18.7%;  Pred. No. 0.033;
Matches  174;  Conservative  116;  Mismatches  330;  Indels  311;  Gaps   38;

Qy         32 PGRQRRVLDPKDAPVQFRVPPSSPACV---SGRAGPHRGNATSFVFKQKTITTWMDTKGP 88
              | :|:|    :  |   |||| |: |    | |          |:               |
Db        540 PPQQQRQKQQRSQPAPARVPPQVPSQVPVTGGVAADEPPPPCSY--------------SP 585

Qy         89 KTAESESK--------------------------ENNNTRIDSMMSSVQKDNFYPHKVE 121
                  ||||                           |   |||   |    |  |:|:
Db        586 VAQSSESKIEPVDVKPRVAPVPPQVPVTPTKPVITNNKKKRID----VVTLDEDAPRRVQ 641

Qy        122 KLENVPQLN-------LDKSPTEKSSQYLNQQQTASVC---------KWQ---------- 155
               : :|:::       | :|| : :  : |: :  |            |::
Db        642 VKQEIPEVSSTSDATKSDAAPTARGAVRIKQEVESDVAPNTILISAKKFERMKAEAEDKE 701

Qy        156 ----------------NEGKHAEQLLASEPPAGTPLPKQ--LSNANIGQ----------- 186
                               | :  |: :|:        :|:    |:  | |
Db        702 DMKKKIAALQEALFNIQEERRVEKEIAAFATTNQAVPQNQPASSVQIAQVSTSESDAPGT 761

Qy        187 ----------SPHTDDH----SDTDHEEDRDNQQFLTPIKLANTK-----------PTVG 221
                        || | ::    :: : |||          ||| :  :              |
Db        762 SEAAATETMTSPKTKNNVIVETEGEQEEDEDE----IPIKKSKKRRAKIVSNDEEEEPVR 817

Qy        222 DGQARSNCKCS----GSRQSVKDCTGCQQEEVDVLPESP------LSDVGAEDIGTGPKN 271
               : ||: :           |  |     ::: :   ||      |:    :|||:   :
Db        818 HPKRRSDEKREKRHVSYAESDDDMPVVKKKRRNQSPEDPEYSAASPSEDEDDDIGSFVVS 877

Qy        272 DNKLTGQES-SLGDSPPFEKESEPESPM-DVDNSRNSCQDSEADEETSPVFDEQDDRSSQ 329
               ||:    :|  :||  | | | | | |  | :  :||    || : :  :|  | :  | :
Db        878 DNEDDDADSFVVGDDEPIEYEEEDEDDMIERRSSRKRRSDSRSKKSATPT-DRRRSRDTP 936

Qy        330 TANKLSSCQAREADGDLRKRYLTKGS-EVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVE 388
               | ::  | ::   :   :|      |: :|      :|   :    |      |  | :||:|:
Db        937 TGSR--SMRSTSPNDRRKSRETPPGNRSMRRTSPSDGRKSRDTP--TASSSMSSSTLSY- 991

Qy        389 CRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERK------------------ 430
               |: ||:       :|          : : :|: |: | :  |||
Db        992 CKKSKETPMSYEEIE------QQKKAKRQRNCKTREENRERKRLAQLEELESSETTGVRR 1045

Qy        431 IPKYIPPNLPPEKKWLGTPIEEMRKMPR-------------------------------C 459
                :     |   |   |  | ||| ||   :
Db       1046 TLRSTQDNSDPLDASLATTIEEFRKTKKKDAKSSENRAKEKQKPMNKRPTSSASVDSNDD 1105

Qy        460 GIHLPSLRPSASHTVTVRVDLLRAGEVPKPF-------------PTHYKDLWDNKHVKMP 506
              |:|:|: |           :  |  ||  |              | |  :: |  :: :
Db       1106 GVHIPAKR------------MAHASSVPGPSRSKPPMIGAVKNRPNH-TEMLDKRNKE-- 1150

Qy        507 CSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNL-----KDAILKYNV------- 554
                ||:     || ||     :     :   :  |  |  ||       | | || |
Db       1151 -SEEKRRKDRDEL-ERLRNKKHTTEEEKI--KMARLQNALKVVGKAAGLKATVKKELTGS 1206

Qy        555 ---------AYSKKWDF-----------TALVDFWDKVLEE------AEAQHLYQSILPD 588
                        |  |  ||           ||:   :| ||             |||     ||
Db       1207 PAKKQKPAPAVPKILDFSVGRTFTAIRQTAIKLVFDTFLERDSPNAAREAQEFELSIAKQ 1266

Qy        589 MVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQIA--SLLANAFFCTFPRRNAKMKSEYS 646
                           ||  |  |: :  : ::: ::  :||               |: |  :
Db       1267 YTDGQKYRINIGHKVAALRKENTSGILEVNKNAVSHDKILAGG-----PKDNCTVARGRK 1321

Qy        647 SYPDINFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQS------------- 693
               :: |               |||  |    |:  |  :  |      |
Db       1322 THVD---------HRQLSIEKLHPLLLQFKLTTSELETNAYPMRRDGSTKAVSIADTVYT 1372

Qy        694 ------LEDFPEWERCEKPLTRLHVTYEGTI 718
                      |:|:    | :    :: ||:
Db       1373 QNKKMFLDDYDMSRNCSRCNKEFKLSPNGTM 1403

RESULT 3
BIR1_SCHPO
ID   BIR1_SCHPO     STANDARD;      PRT;   997 AA.
AC   O14064; Q9USG4;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Bir1 protein (Chromosome segregation protein cut17).
GN   BIR1 OR CUT17 OR PBH1 OR SPCC962.02C OR SPCP31B10.10C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21439264; PubMed=11554922;
RA   Morishita J., Matsusaka T., Goshima G., Nakamura T., Tatebe H.,
RA   Yanagida M.;
RT   "Bir1/Cut17 moving from chromosome to spindle upon the loss of
RT   cohesion is required for condensation, spindle elongation and
RT   repair.";
RL   Genes Cells 6:743-763(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99398681; PubMed=10468581;
RA   Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,
RA   Vaux D.L., Lithgow T.;
RT   "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell
RT   division.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175(1999).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=21850422; PubMed=11861551;
RA   Rajagopalan S., Balasubramanian M.K.;
RT   "Schizosaccharomyces pombe Bir1p, a nuclear protein that localizes to
RT   kinetochores and the spindle midzone, is essential for chromosome
RT   condensation and spindle elongation during mitosis.";
RL   Genetics 160:445-456(2002).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=20035862; PubMed=10571085;
RA   Rajagopalan S., Balasubramanian M.K.;
RT   "S. pombe Pbh1p: an inhibitor of apoptosis domain containing protein
RT   is essential for chromosome segregation.";
RL   FEBS Lett. 460:187-190(1999).
CC   -!- FUNCTION: Seems to act in the pleiotropic control of cell
CC       division. Has a role in chromosome segregation by recruiting
```

```
CC       condensin and ark1 kinase to appropriate sites as the cell
CC       progresses through mitosis.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Interacts with the outer
CC       centromeric regions of the chromosomes during interphase. After
CC       chromatid separation moves to the middle of the spindle.
CC   -!- SIMILARITY: Contains 2 BIR repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB031034; BAA83415.1; -.
DR   EMBL; AL031323; CAA20434.1; -.
DR   EMBL; AL121859; CAB58376.1; -.
DR   PIR; T43523; T43523.
DR   HSSP; Q13490; 1QBH.
DR   GeneDB_SPombe; SPCC962.02c; -.
DR   InterPro; IPR001370; BIR.
DR   Pfam; PF00653; BIR; 2.
DR   SMART; SM00238; BIR; 2.
DR   PROSITE; PS01282; BIR_REPEAT_1; FALSE_NEG.
DR   PROSITE; PS50143; BIR_REPEAT_2; 2.
KW   Cell division; Mitosis; Nuclear protein; Repeat.
FT   REPEAT       25     99       BIR 1.
FT   REPEAT      120    194       BIR 2.
FT   DOMAIN       80     83       POLY-ASP.
FT   DOMAIN      312    319       POLY-ASP.
FT   DOMAIN      487    490       POLY-SER.
SQ   SEQUENCE   997 AA;  112579 MW;  952A6BAFA5C489F4 CRC64;

  Query Match             3.3%;  Score 170;  DB 1;  Length 997;
  Best Local Similarity   21.2%;  Pred. No. 0.021;
  Matches  135;  Conservative   82;  Mismatches  227;  Indels  194;  Gaps   32;

Qy          7 WEP--------------CTKARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPP 52
Db        173 WEPDDDPYTEHKRRRADCVFFTWKDPNSLSPT---KLSFLSTSN--IDPEDLTEDNSILP 227

Qy         53 SSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAE-SESKENN-NTRIDSMM--- 107
Db        228 VSP------------------------TRDSTKSHKTLNFSPSRKNNLNARPLTMSLYT 262

Qy        108 -SSVQKDNFYPHKVEKLENVPQL-----NLDKSPTEKSSQYLNQQQTASVCK-----WQN 156
Db        263 NTSEEKDS-QPTRAPQSPTKPVLLTAPRRKNKSP---------KKSKPAVFKPVKPIFSD 312

Qy        157 EGKHAEQLLASEPPAGTPLPKQLSN--ANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLA 214
Db        313 EDEDDDDLTASQ-----PFSKGICNDSMQVAKKNFTEEIPLKEDEKDNELEHLVSPATSV 367

Qy        215 NTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNK 274
Db        368 HT--------------------TVSDITGHQS----------VTDESDE----------- 386

Qy        275 LTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSSQTANKL 334
Db        387 ---QNNCMSTPPKIEIESKIEEEISVVS-----KSKEISSSVSSVGKEQNHTEKQVAIET 438

Qy        335 SSCQAREADGDLRKRYLTKGSEVRLHFQFEGE--NNAGTSDLNAKPSGNSSSLNVECRSS 392
Db        439 PEQQKVEKEDE--------------HLNLQGSFIEESTKQPISSKPSTSSPDMTDAATGG 484

Qy        393 KQHGK--RDS---------KITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPP 441
Db        485 RVSSSSFRDKILQTNFSPRSTIDSFSNISKK--RNSEEANDENDETNLKIP-------IP 535

Qy        442 EKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNK 501
Db        536 EKKRKFQEVLQSKNI------LVSSTEDSHEPVKVTEDSQTAIHVSK-----FEDL-ENK 583

Qy        502 HVKMPCSEQNLYPV-EDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS--- 557
Db        584 SME---SEQSLQLLSESENDDK------PLIDLIPLLAIKRKDNLVSGVLEKGKSTSTSK 634

Qy        558 KKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALC 595
Db        635 TKFD-TSIVDFIEKPKTE------ISEVLPEEKRKAIC 665


RESULT 4
AF4_HUMAN
ID   AF4_HUMAN      STANDARD;      PRT;  1210 AA.
AC   P51825;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   AF-4 protein (Proto-oncogene AF4) (FEL protein).
GN   MLLT2 OR AF4 OR FEL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281633; PubMed=8506309;
RA   Nakamura T., Alder H., Gu Y., Prasad R., Canaani O., Kamada N.,
RA   Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
RA   Canaani E.;
RT   "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT   acute leukemia share sequence homology and/or common motifs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93184301; PubMed=8443374;
RA   Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
RA   Cleary M.L.;
RT   "A serine/proline-rich protein is fused to HRX in t(4;11) acute
RT   leukemias.";
RL   Blood 81:1124-1131(1993).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DISEASE: Involved in acute leukemias through a chromosomal
CC       translocation t(4;11)(q21;q23) that involves MLLT2 and MLL/HRX.
CC       The result is a rogue activator protein.
CC   -!- SIMILARITY: Belongs to the AF4 family.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/AF4.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L13773; AAA58360.1; -.
DR   EMBL; L25050; AAA36642.1; -.
DR   PIR; A58198; A58198.
DR   PIR; I39410; I39410.
DR   Genew; HGNC:7135; MLLT2.
DR   MIM; 159557; -.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   InterPro; IPR007797; AF-4.
DR   Pfam; PF05110; AF-4; 1.
KW   Nuclear protein; Chromosomal translocation; Proto-oncogene.
FT   DOMAIN      483    492       POLY-SER.
FT   DOMAIN      835    843       POLY-SER.
FT   DOMAIN      866    869       POLY-PRO.
FT   DOMAIN      871    874       POLY-SER.
FT   CONFLICT     46     46       K -> R (IN REF. 2).
FT   CONFLICT    624    624       E -> G (IN REF. 2).
FT   CONFLICT    899    905       SASSTKS -> VPAVPRV (IN REF. 2).
```

```
FT   CONFLICT     928    929       EH -> AD (IN REF. 2).
FT   CONFLICT     999    999       I -> N (IN REF. 2).
FT   CONFLICT    1096   1096       A -> AR (IN REF. 2).
FT   CONFLICT    1140   1140       N -> I (IN REF. 2).
FT   CONFLICT    1177   1210       STNVCTLALNSSLVDLVHYTRQGFQQLQELTKTP -> RQ
FT                                 MCAPWPSTAVWWTWCTIHDRVFSSYKN (IN REF. 2).
SQ   SEQUENCE   1210 AA;  131421 MW;  F0E334DF8FC2FF04 CRC64;

  Query Match              3.2%;  Score 163.5;  DB 1;  Length 1210;
  Best Local Similarity   19.7%;  Pred. No. 0.065;
  Matches  197;  Conservative  112;  Mismatches  368;  Indels  321;  Gaps   42;

Qy         35 QRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE 94
Db        260 QDSLVAPAQPPSQTFPPPSLP-------------SKSVAMQQKPTAYVRPMDGQDQAPSE 306

Qy         95 SKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKW 154
Db        307 SPELKPLPEDYRQQTFEKTDL---KVPAKAKLTKLKMPSQSVEQT-------------Y 349

Qy        155 QNEGKHAEQLL-----------------ASEPPAGTPLP-------------------- 176
Db        350 SNEVHCVEEILKEMTHSWPPPLTAIHTPSTAEPSKFPFPTKDSQHVSSVTQNQKQYDTSS 409

Qy        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIK--LANTKPTV------------G 221
Db        410 KTHSNSQQGTSSMLEDDLQLSDSEDSDSEQ--TPEKPPSSSAPPSAPQSLPEPVASAHSS 467

Qy        222 DGQARSNCKCSGSRQSVKDCTGCQQEEVDVL------PESPLSDVGAEDIGTGPKNDNKL 275
Db        468 SAESESTSDSDSSSDSESESSSSDSEENEPLETPAPEPEPPTTNKW--------QLDNWL 519

Qy        276 TGQESSLGDSPPFEKESEPESPMDVDNS-----RNSCQDSEADEETSPVFDEQDDRSSQT 330
Db        520 T-------------KVSQPAAPPEGPRSTEPPRRHPESKGSSDSATSQEHSESKDPPPKS 566

Qy        331 ANKL------------SSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKP 378
Db        567 SSKAPRAPPEAPHPGKRSCQKSPAQQEPPQR-----------------QTVGTKQ-PKKP 608

Qy        379 SGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPN 438
Db        609 V--KASARAGSRTSLQ-GEREPGLLPYGSRDQTSKDKPKVKTKGRPRAAASNEPK---PA 662

Qy        439 LPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLW 498
Db        663 VPPSSE---------KKKHKSSLPAPS----------------KALSGPEPAKDNVEDRT 697

Qy        499 DNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSK 558
Db        698 PEHFALVPLTESQ-GPPHSGSGSRTSGCRQAVVVQEDSRKDRLPLPLRDT---------- 746

Qy        559 KWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQP--------------- 603
Db        747 ------------KLLSPLRDTPPPQSL---MVKITLDLLSRIPQPPGKGSRQRKAEDKQP 791

Qy        604 ----------------IPLLKQKM--------NHSVTM-----SQEQIASLLANAFFCTF 634
Db        792 PAGKKHSSEKRSSDSSSKLAKKRKGEAERDCDNKKIRLEKEIKSQSSSSSSHKESSKTK 851

Qy        635 PRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEK--LKTLFCYFRRVTE---KKPTGLVTF 689
Db        852 PSRPSSQSSKKEMLPPPPVS-----SSSQKPAKPALKR----SRREADTCGQDPPKSASS 902

Qy        690 TRQSLED--FPEWERCEKPLTRLHVTYEGTI-------------EGNGR-GMLQVDFANR 733
Db        903 TKSNHKDSSIPKQRRVEGKGSRSSSEHKGSSGDTANPFPVPSLPNGNSKPGKPQVKFDKQ 962

Qy        734 FVGGGVTGAGL-VQEEIRFLINPELIVSRL------FTEVLDHNECLIITGTEQYSE--- 783
Db        963 -------QADLHMREAKKMKQKAELMTDRVGKAFKYLEAVLSFIECGIATESESQSSKSA 1015

Qy        784 YTGYAET------YRWARSHEDGSEKDDWQ------RRCTEIVAIDALHFRRYLDQFVPE 831
Db       1016 YSVYSETVDLIKFIMSLKSFSDATAPTQEKIFAVLCMRCQSI--LNMAMFRCKKD--IAI 1071

Qy        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGG 869
Db       1072 KYSRTLNKHF----------ESSSKVAQAPSPCIASTG 1099

RESULT 5
PCLO_MOUSE
ID   PCLO_MOUSE     STANDARD;      PRT;  5038 AA.
AC   Q9QYX7; Q9QYX6; Q9QZJ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-
DE   derived HLMN protein).
GN   PCLO OR ACZ.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RP   TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   REVISIONS.
RA   Kilimann M.W.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 4502-4682 FROM N.A.
RC   TISSUE=Brain;
RA   Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   INTERACTION WITH RIMS2.
RX   MEDLINE=22384373; PubMed=12401793;
RA   Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA   Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT   "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT   cAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis.";
RL   J. Biol. Chem. 277:50497-50502(2002).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking.
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin.
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9QYX7-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC       stomach. Not detected in other tissues analyzed including adrenal
CC       gland, testis and pancreas.
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Y19185; CAB60731.2; -.
DR   EMBL; Y19186; CAB60732.2; -.
DR   EMBL; AF181269; AAD55786.2; -.
DR   HSSP; P04410; 1A25.
DR   MGD; MGI:1349390; Pclo.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; IDA.
DR   GO; GO:0019933; P:cAMP-mediated signaling; IDA.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
DR   GO; GO:0030073; P:insulin secretion; IDA.
DR   GO; GO:0017157; P:regulation of exocytosis; IDA.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      371    470       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     502    526       C4-TYPE (POTENTIAL).
FT   ZN_FING     967    990       C4-TYPE (POTENTIAL).
FT   DOMAIN     2305   2329       POLY-PRO.
FT   DOMAIN     4394   4488       PDZ.
FT   DOMAIN     4607   4705       C2 DOMAIN 1.
FT   DOMAIN     4922   5012       C2 DOMAIN 2.
FT   VARSPLIC   4829   4833       TKPTN -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003928.
FT   VARSPLIC   4834   5038       Missing (in isoform 2).
FT                                /FTId=VSP_003929.
SQ   SEQUENCE   5038 AA;  547600 MW;  DADA460CF3B40888 CRC64;

  Query Match              3.1%;  Score 162;  DB 1;  Length 5038;
  Best Local Similarity   19.9%;  Pred. No. 0.52;
  Matches  122;  Conservative   86;  Mismatches  207;  Indels  198;  Gaps   28;

Qy         75 KQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKS 134

Db       1355 QQKTQTP-SETRDISISEEEIKESQEKKVTSKKDSAQG---FPSRKEHKEN-PELVDDLS 1409

Qy        135 PTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHS 194

Db       1410 PRRASYDSVED---------------------SSESENSPVARRKRRTSIGSS-SSEEYK 1447

Qy        195 DTDHEEDRDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSV------KDCTGCQQEE 248

Db       1448 QEDSQGSGEDEDFI--------------------------RKQIIEMSADEDASGSEDEE 1481

Qy        249 VDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQ 308

Db       1482 ---FIRSQLKEIGG--VTESQKREETKGKGKSPAGKHRRLTRKS--STSFDDDAGRRHSW 1534

Qy        309 DSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRK------------RYLTKGS 355

Db       1535 HDEDDE----TFDESPELKFRETKSQESEELVVAGGGGLRRFKTIELNSTVTDKYSAESS 1590

Qy        356 EVR--LHFQFEGE--------------NNAGTSDLNAK--PSGNS----SSLNVECRSSK 393

Db       1591 QKKTTLYFDEEPELEMESLTDSPEDRSRGEGSSSLHASSFTPGTSPTSVSSLDEDSDSSP 1650

Qy        394 QHGKRDSK----------------ITD-----------------------HFMRISKSE 413

Db       1651 SHKKGESKQQRKARHRSHGPLLPTIEDSSEEEELREEEELLKEQEKQRELEQQQRKSSSK 1710

Qy        414 DRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPR--------CGIHLPS 465

Db       1711 KSKKDKDELRAQR-RRERPKTPPSNLSPIED--ASPTEELRQAAEMEELHRSSCSEYSPS 1767

Qy        466 LRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWD-NKHVKMPCSEQNLYPVEDENGERTA 524

Db       1768 IESD-----------------PEGFEISPEKIIEVQKVYKLP-TAVSLYSPTDEQ----- 1804

Qy        525 GSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTAL--VDFWDKVLEEAE----- 577

Db       1805 ---------SVMQKEGAQKALKSAEEMYEEMMHKPHKYKAFPAANERDEVFEKEPLYGGM 1855

Qy        578 --AQHLYQSILPD 588

Db       1856 LIEDYIYESLVED 1868

RESULT 6
PCLO_HUMAN
ID   PCLO_HUMAN     STANDARD;      PRT;  5147 AA.
AC   Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragments).
GN   PCLO OR ACZ OR KIAA0559.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
```

```
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 4405-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9Y6V0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC                                  VSP_003926, VSP_003927;
CC         Note=No experimental confirmation available;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19188; CAB60727.1; -.
DR   EMBL; AC004903; AAD20936.1; -.
DR   EMBL; AC004886; AAD21789.1; -.
DR   EMBL; AB011131; BAA25485.1; -.
DR   EMBL; BC001304; AAH01304.1; -.
DR   EMBL; AC004082; AAB97937.1; -.
DR   PIR; T00634; T00634.
DR   HSSP; P04410; 1A25.
DR   Genew; HGNC:13406; PCLO.
DR   MIM; 604918; -.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0045202; C:synaptic junction; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
KW   Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN      400    465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     499    523       C4-TYPE (POTENTIAL).
FT   ZN_FING     969    992       C4-TYPE (POTENTIAL).
FT   NON_CONS   1010   1011
FT   DOMAIN     2300   2325       POLY-PRO.
FT   DOMAIN     4391   4442       PDZ.
FT   DOMAIN     4544   4633       C2 DOMAIN 1.
FT   DOMAIN     5031   5121       C2 DOMAIN 2.
FT   VARSPLIC   4404   4404       S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                                QTGKLMEG (in isoform 2).
FT                                /FTId=VSP_003923.
FT   VARSPLIC   4534   4534       K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
FT                                /FTId=VSP_003924.
FT   VARSPLIC   4576   4576       G -> GQVMVVQNAS (in isoform 2).
FT                                /FTId=VSP_003925.
FT   VARSPLIC   4757   4761       TAHKS -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003926.
FT   VARSPLIC   4762   5147       Missing (in isoform 2).
FT                                /FTId=VSP_003927.
SQ   SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

  Query Match             3.1%;  Score 161;  DB 1;  Length 5147;
  Best Local Similarity   18.7%;  Pred. No. 0.61;
  Matches  114;  Conservative   85;  Mismatches  189;  Indels  222;  Gaps   25;

Qy         91 AESESKENNNTRIDSMMSSVQKD--NFYPHKVEKLENVPQLNLDKSP---TEKSSQYLNQ 145
Db       1359 SEEEIKESQEERKDTFKKDSQQDIPSSKDHK-EKSEFVDDITTRREPYDSVEESSESEN- 1416

Qy        146 QQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHSDTDHEEDRDNQ 205
Db       1417 -------------------------SPVPQRKRRTSVGS-------SSSDEYKQEDSQ 1442

Qy        206 QFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDI 265
Db       1443 ---------------GSG----------------------EEEDFI-RKQIIEMSADED 1463

Qy        266 GTGPKNDNKLTGQESSLGDSPPFEKESEPE-----------------SPMDVDNSRNSC 307
Db       1464 ASGSEDDEFIRNQLKEISSSTESQKKEETKGKGKITAGKHRRLTRKSSTSIDEDAGRRHS 1523

Qy        308 QDSEADE--ETSPVFDEQDDRSSQT-------ANKLSSCQAREADGDLRKRYLTKGSEVR 358
Db       1524 WHDEDDEAFDESPELKYRETKSQESEELVVTGGGGLRRFKTIELNSTIADKYSAESSQKK 1583

Qy        359 LHFQFEGE----------------NNAGTSDLNAK--PSGNS----SSLNVECRSSKQHG 396
Db       1584 TSLYFDEEPELEMESLTDSPEDRSRGEGSSSLHASSFTPGTSPTSVSSLDEDSDSSPSHK 1643

Qy        397 KRDSK----------------ITD-----------------------HFMRISKSEDRR 416
Db       1644 KGESKQQRKARHRPHGPLLPTIEDSSEEEELREEEELLKEQEKQREIEQQQRKSSSKKSK 1703

Qy        417 KEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPR--------CGIHLPSLRP 468
Db       1704 KDKDELRAQR-RRERPKTPPSNLSPIED--ASPTEELRQAAEMEELHRSSCSEYSPSIES 1760

Qy        469 SASHTVTVRVDLLRAGEVPKPFPTHYKDLWD-NKHVKMPCSEQNLYPVEDENGERTAGSR 527
Db       1761 D----------------PEGFEISPEKIIEVQKVYKLP-TAVSLYSPTDEQ-------- 1794

Qy        528 WELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTAL--VDFWDKVLEEAE-------A 578
Db       1795 ------SIMQKEGSQKALKSAEEMYEEMMHKTHKYKAFPAANERDEVFEKEPLYGGMLIE 1848

Qy        579 QHLYQSILPD 588
Db       1849 DYIYESLVED 1858


RESULT 7
DSPP_RAT
ID   DSPP_RAT      STANDARD;      PRT;   687 AA.
```

```
AC   Q62598; P70578; Q9R057;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin sialophosphoprotein precursor [Contains: Dentin phosphoprotein
DE   (Dentin phosphophoryn) (DPP); Dentin sialoprotein (DSP)].
GN   DSPP OR RDSP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM DPP-2).
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20435277; PubMed=10978503;
RA   Ritchie H.H., Wang L.-H.;
RT   "The presence of multiple rat DSP-PP transcripts.";
RL   Biochim. Biophys. Acta 1493:27-32(2000).
RN   [2]
RP   SEQUENCE OF 1-387 FROM N.A., AND SEQUENCE OF 18-26.
RC   STRAIN=Sprague-Dawley; TISSUE=Odontoblast;
RX   MEDLINE=94148875; PubMed=8106414;
RA   Ritchie H.H., Hou H., Veis A., Butler W.T.;
RT   "Cloning and sequence determination of rat dentin sialoprotein, a
RT   novel dentin protein.";
RL   J. Biol. Chem. 269:3698-3702(1994).
RN   [3]
RP   SEQUENCE OF 421-687 FROM N.A. (ISOFORM DPP-1).
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=96355551; PubMed=8702961;
RA   Ritchie H.H., Wang L.-H.;
RT   "Sequence determination of an extremely acidic rat dentin
RT   phosphoprotein.";
RL   J. Biol. Chem. 271:21695-21698(1996).
RN   [4]
RP   SEQUENCE OF 1-19 FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=99333695; PubMed=10403786;
RA   Yamazaki H., Kunisada T., Miyamoto A., Tagaya H., Hayashi S.-I.;
RT   "Tooth-specific expression conferred by the regulatory sequences of
RT   rat dentin sialoprotein gene in transgenic mice.";
RL   Biochem. Biophys. Res. Commun. 260:433-440(1999).
RN   [5]
RP   SEQUENCE OF 29-33; 70-79; 93-109; 136-148; 162-188; 266-308; 398-423
RP   AND 426-438, AND PHOSPHORYLATION OF SER-292 AND SER-298.
RX   MEDLINE=21125612; PubMed=11042175;
RA   Qin C., Cook R.G., Orkiszewski R.S., Butler W.T.;
RT   "Identification and characterization of the carboxyl-terminal region
RT   of rat dentin sialoprotein.";
RL   J. Biol. Chem. 276:904-909(2001).
RN   [6]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=98055479; PubMed=9395101;
RA   Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
RA   Hotton D., Papagerakis P., Berdal A., Butler W.T.;
RT   "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
RT   expression in odontoblasts and transient expression in
RT   pre-ameloblasts.";
RL   Eur. J. Oral Sci. 105:405-413(1997).
CC   -!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
CC       may bind high amount of calcium and facilitate initial
CC       mineralization of dentin matrix collagen as well as regulate the
CC       size and shape of the crystals.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPP-1; Synonyms=PP240;
CC         IsoId=Q62598-1; Sequence=Displayed;
CC       Name=DPP-2; Synonyms=PP171;
CC         IsoId=Q62598-2; Sequence=VSP_003855;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Specifically expressed in teeth, mainly in
CC       odontoblasts and transiently in pre-ameloblasts.
CC   -!- PTM: DSP is glycosylated.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 380.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF247187; AAK96895.1; -.
DR   EMBL; U02074; AAA18932.1; ALT_FRAME.
DR   EMBL; U63111; AAC52774.1; -.
DR   EMBL; AF114987; AAD48588.1; ALT_TERM.
KW   Biomineralization; Extracellular matrix; Signal; Glycoprotein;
KW   Phosphorylation; Sialic acid; Alternative splicing.
FT   SIGNAL        1     17
FT   CHAIN        18    687       DENTIN SIALOPHOSPHOPROTEIN.
FT   CHAIN        18    447       DENTIN SIALOPROTEIN.
FT   CHAIN       448    687       DENTIN PHOSPHOPROTEIN.
FT   MOD_RES      57     57       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     226    226       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     253    253       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     278    278       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     292    292       PHOSPHORYLATION (BY CK2).
FT   MOD_RES     298    298       PHOSPHORYLATION (BY CK1).
FT   MOD_RES     315    315       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     319    319       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     329    329       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     337    337       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     345    345       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     366    366       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   CARBOHYD     55     55       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    189    189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    567    635       Missing (in isoform DPP-2).
FT                                /FTId=VSP_003855.
FT   CONFLICT     74     74       N -> D (IN REF. 5; AA SEQUENCE).
FT   CONFLICT    564    564       S -> T (IN REF. 3).
SQ   SEQUENCE   687 AA;  70179 MW;  9A845EED6AA31B63 CRC64;

  Query Match              3.1%;  Score 159;  DB 1;  Length 687;
  Best Local Similarity   20.3%;  Pred. No. 0.056;
  Matches   85;  Conservative   61;  Mismatches  195;  Indels   78;  Gaps   13;

Qy       16 GAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSGRA--------GPHRG 67
            |   ||:   :      | |    :   :| :  :   : || ||::        |   |
Db      309 GGDNTSSSEETGIEEGDGTQTTQDNQNLSPTEGGIISQAEACPSGQSQNQGLETEGSSTG 368

Qy       68 NATSFVFKQKTITTWMDTKG-----------PKTAESE--------SKENNN---TRIDS 105
            | :|   :   ::   |: |           ||  ||:        |  :||   :|| |
Db      369 NKSSITKESGKLSGSKDSNGHHGMELDKRNSPKQGESDKPQGAAEKSDTHNNMGHSRIGS 428

Qy      106 MMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLL 165
              :|   |: |    | ::     : |:|     |   | :           | | :
Db      429 SSNSDGHDS-YDFDDESMQGDDPNSSDESNGSDGSDDANSESAI------ENGNHGDASY 481

Qy      166 ASEPPAGTPLPKQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQA 225
             |:  :        |::: |:   :|| ||||  :   :    :  |         |
Db      482 TSDESSDNGSD---SDSHAGEDDSSDDTSDTDDSDSNGDDDSESKDK---------DESD 529

Qy      226 RSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDS 285
             ||       :|  | :    :  |    |  ||  :|   :   :|   :   |   ||
Db      530 NSNHDNDSDSESKSDSSDSDSDSSDSSDSSD-SSETSDSSDSSDTSDSSDSSDSSDS 588

Qy      286 PPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGD 345
                  |:     | |:|     ||:: :  |       |  |  ::  ||  : ::| |
Db      589 SNSSDTSDSSDSSDGDSSDGDSSDSDSSDSDS---SNSSDSDSSDSSDSSSSDSSDSDSD 645
```

```
Qy        346 LRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITD 404
                :                       :: : :|| |:| ||| :|         |         ||: :|
Db        646 SK-----------------DSTSDSSDDNSK-SGNGNS------DSDSDSDSDSEGSD 679
```

```
RESULT 8
NKCR_HUMAN
ID   NKCR_HUMAN     STANDARD;      PRT;  1462 AA.
AC   P30414;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE   related protein) (NK-TR protein).
GN   NKTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93133824; PubMed=8421688;
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS.
RA   Anderson S.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Component of a putative tumor-recognition complex.
CC       Involved in the function of NK cells.
CC   -!- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC       via its N-terminus.
CC   -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04288; AAA35734.2; -.
DR   EMBL; AF184110; AAD56402.1; -.
DR   PIR; A47328; A47328.
DR   HSSP; Q27450; 1A33.
DR   Genew; HGNC:7833; NKTR.
DR   MIM; 161565; -.
DR   GO; GO:0016018; F:cyclosporin A binding; TAS.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
FT   DOMAIN        1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN      219    240       ARG/LYS-RICH (BASIC).
FT   DOMAIN      421    457       ARG/LYS-RICH (BASIC).
FT   DOMAIN      970   1010       ARG/LYS-RICH (BASIC).
FT   DOMAIN      194    244       ARG/SER-RICH.
FT   DOMAIN      466    574       ARG/SER-RICH.
FT   DOMAIN      664    814       ARG/SER-RICH.
FT   DOMAIN     1311   1348       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE   1462 AA;  165676 MW;  D98A1147763EF527 CRC64;

  Query Match             3.1%;  Score 159;  DB 1;  Length 1462;
  Best Local Similarity   19.2%;  Pred. No. 0.15;
  Matches  146;  Conservative   99;  Mismatches  252;  Indels  262;  Gaps   32;
```

```
Qy         20 TSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTI 79
              |:: ::| |||                          | :  | ::| || |
Db        537 TASKSSSHSRS---------------------RSKSRSSSKSG-HRKRA---------- 563

Qy         80 TTWMDTKGPKTAESESKENNNTRIDSMMSSV-QKDNFYPHKVEKLENVPQLNLDKSP--- 135
                   :| |:   |:  ||   : : : ::: | :|     |    ||:| : |  ||
Db        564 -----SKSPRKTASQLSENKPVKTEPLRATMAQNENVVVQPV-VAENIPVIPLSDSPPPS 617

Qy        136 -------------------TEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176
                                 |::  |  | | |:  :  |              : :
Db        618 RWKPGQKPWKPSYERIQEMKAKTTHLLPIQSTYSLANIKETG------------SSSSYH 665

Qy        177 KQLSNANIGQSPHT--DDHSDTDHEEDRD--------NQQFLTPIKLANTKPTVGDGQAR 226
              |:  |:   || ::   | |         |           :: :     ||::      ::|
Db        666 KREKNSESDQSTYSKYSDRSSESSPRSRSRSSRSRSYSRSYTRSRSLASS-----HSRSR 720

Qy        227 SNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGD-- 284
              |    | ||     | : | :          |  : : ::| |     |     : :|:::|:
Db        721 SPSSRSHSRNKYSDHSQCSR-------SSSYTSISSDD-GRRAKRRLRSSGKKNSVSHKK 772

Qy        285 -------------------SPPFEKESEPESPMDV--DNSRNSCQDSEADEETSPVFDEQ 323
                                 |     | ||  | :|   |: ::| | ::: :|       |:
Db        773 HSSSSEKTLHSKYVKGRDRSSCVRKYSESRSSLDYSSDSEQSSVQATQSAQE-----KEK 827

Qy        324 DDRSSQTANKLS-------------------------SCQAREADGDLRKRY------- 350
                :  :| ||                                  |  |     || |
Db        828 QGQMERTHNKQEKNRGEEKSKSERECPHSKKRTLKENLSDHLRNGSKPKRKNYAGSKWDS 887

Qy        351 -------LTKGSEVRLHFQFEGENNAGTSD----------LNAKPSGNSS---SL---NVE 388
                     :|| |:   |    :  |    |||            :  ||:  ||   ||   |
Db        888 ESNSERDVTKNSKNDSHPSSDKEEGEATSDSESEVSEIHIKVKPTTKSSTNTSLPDDNGA 947

Qy        389 CRSSKQHGKRDSKITDHFMRISKSEDRRKEQCEVRH---QRTERKIPKYIPPNLPPEKKW 445
                :||||     :  :|      | ||: | :  : :|    :  :|:  |  :   |  :|
Db        948 WKSSKQR----TSTSDSEGSCSNSENNRGKPQKHKHGSKENLKREHTKKVKEKLKGKK-- 1001

Qy        446 LGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKM 505
                    ::  | |:              |  | |                : |  |    ::  |:| |
Db       1002 -----DKKHKAPK---------RKQAFH-----------WQPPLEFGEEEEEEIDDKQVTQ 1037

Qy        506 PCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVA----YSKKWD 561
              |: :        || |                       :|| |||   :       | | |
Db       1038 ESKEKKV----SENNE---------------------TIKDNILKTEKSSEEDLSGKHD 1071

Qy        562 FTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLKQKMNHSVTMSQEQ 621
                  :      |: ::       :  : |       :| || ||              | :  ||     |
Db       1072 TVTVSSDLDQFTKDDSKLSISPTALNTEENVA-CLQNI---------QHVEESVPNGVED 1121

Qy        622 IASLLANAFFCTFPRRNAKMKSEYSS--------YPDIN 652
                :      |     || |  |:: |  | :|                ||||
Db       1122 VLQTDDNMEICT-PDRSSPAKVEETSPLGNARLDTPDIN 1159
```

```
RESULT 9
FA5_BOVIN
ID   FA5_BOVIN      STANDARD;      PRT;  2211 AA.
AC   Q28107; Q28108;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Coagulation factor V precursor (Activated protein C cofactor).
GN   F5.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92147638; PubMed=1737753;
RA   Guinto E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19187; CAB60725.1; -.
DR   HSSP; P04410; 1A25.
DR   GO; GO:0045202; C:synaptic junction; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001565; Synaptotagmin.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat.
FT   NON_TER       1      1
FT   DOMAIN      258    357       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     368    392       C4-TYPE (POTENTIAL).
FT   ZN_FING     836    859       C4-TYPE (POTENTIAL).
FT   DOMAIN     2324   2343       POLY-PRO.
FT   DOMAIN     4414   4493       PDZ.
FT   DOMAIN     4627   4726       C2 DOMAIN 1.
FT   DOMAIN     5003   5094       C2 DOMAIN 2.
SQ   SEQUENCE   5120 AA;  560751 MW;  A658D9891B65B412 CRC64;

  Query Match             3.0%;  Score 152.5;  DB 1;  Length 5120;
  Best Local Similarity   17.9%;  Pred. No. 1.9;
  Matches  115;  Conservative  122;  Mismatches  253;  Indels  151;  Gaps   29;

Qy         40 DPKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENN 99
              | || | :       ||     :| :  |    :          : |: : :: :      : :| :
Db       1302 DEKDLPRE----PSQKDTISQESPPSPSDLAKL---ESTVLSILEAQASTLTDEKSVKRK 1354

Qy        100 NTRIDSMMSSVQKDNFYPHKVEKLENVPQ-LNLDKSPTEKSSQYLNQQQTASVCKW--QN 156
              :     |  ||   || :   |    |: :  |:  |       :: |:    |          :
Db       1355 --ELYETYSEQTKDQ---HKTKPLPVTPESYSSDEEDLEAIQE--GERTIAADSKGGASS 1407

Qy        157 EGKHAEQLLASEPPA---------------GTPLPKQLSNANIGQSPHTDDHSDTDHEED 201
              :  : |:    :: ||                :|:|::    |::|  |   :|::     |  :
Db       1408 QTDYKEEDGGNDTPARRQRYDSVEDSSESENSPVPRRKRRASVGSS-SSDEYKRDDSQGS 1466

Qy        202 RDNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVG 261
              | : |:          :   :  :  |    :    |     : :::      | :|: |        :|
Db       1467 GDEEDFI--------RKQIIEMSADEDASGSEDDEFIRN----QLKEISVTESQKKEEVK 1514

Qy        262 AEDIGTGPKNDNKLTGQESSLG--------------DSPPFEKESEPE----SPMDVD-- 301
              ::   ||   |   ::    ::|| |                  | |:  :  ||:         | :
Db       1515 SKAKGTVGK--HRRMARKSSAGYDEDAGRRHSWHDDDDETFDESPEPKYRETKSQDGEEL 1572

Qy        302 ----------------NSRNSCQDSEADEETSPV--FDEQDDRSSQTANKLSSCQAR-EA 342
                              ||  :  :  || |:    :   |||:  :     ::         ::| |
Db       1573 AISGGGGLRRFKTIELNSTITSKYSETPEQQKGILYFDEEPELEMESLTDSPEDRSRGEG 1632

Qy        343 DGDLRKRYLTKGSEVRLHFQFEGENNA--------GTSDLNAK---------PSGNSSS 384
                   |      | |:            : ::::           | |  |              |:  ||
Db       1633 SSSLHASSFTPGTSPTSVSSLDEDSDSSPSHKKLGGESKQQRKARHRSHGPLLPTIEDSS 1692

Qy        385 LNVECRSSKQHGKRDSK--ITDHFMRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPPE 442
                   | |  ::  |   |    :    | | |:  :|::  |:| ||   |: ||   | || |
Db       1693 EEEELREEEELLKEQEKQRELEQQQRKSSSKKSKKDKDELRAQR-RRERPKTPPSNLSPI 1751

Qy        443 KKWLGTPIEEMRKMPR--------CGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHY 494
                :    :| ||:|:               |   :  ||:               ::  :|  |  ||
Db       1752 ED--ASPTEELRQAAEMEELHRSSCSEYSPSIESDPEGFEISPEKIIEVQKVYK-LPTAV 1808

Qy        495 KDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNV 554
                          :  |   |: :   :::|:|::|   |   |:  :   :::|   : ::          : :
Db       1809 S-------LYSPTDEKLIGALKEESGQKTLKSAEEVYE-EMIHKTHKSKS-------FQI 1853

Qy        555 AYSKKWDFTALVDFWDKVLEEAE-------AQHLYQSILPD 588
              |  |              |:| |:             ::|:|:: |
Db       1854 ASEK-----------DEVFEKESLYGGMLIEDYIYESLIED 1883


RESULT 15
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   PIR; T34036; T34036.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;
KW   DNA-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
```

```
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;

  Query Match             2.9%;  Score 152;  DB 1;  Length 1359;
  Best Local Similarity  19.0%;  Pred. No. 0.35;
  Matches  141;  Conservative  111;  Mismatches  275;  Indels  214;  Gaps   32;

Qy         86 KGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLNQ 145
              |  | |:|||: : :    :        |   |       | || :    :     |  :: |     :
Db         81 KSRKRAKSESESDESDEEEDRKKSKSKKKVDQKKKEKSK---KKRTTSSSEDEDSDEERE 137

Qy        146 QQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHSDTDHEEDRDNQ 205
              |::      |    :   :|     ||         |:     :: :    | :  || |  :   : ::
Db        138 QKSKKKSKKTKKQTSSESSEESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSKKSK 197

Qy        206 QFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAEDI 265
              : |     | | ::     : ::     :     |::  |       : |   |   ||
Db        198 KGLK--KKAKSE---SESESEDEKEVKKSKKKSKKVVKKESESEDEAPEKK--------- 243

Qy        266 GTGPKNDNKLTGQESS---LGDSPPFEKESEP----ESPMDVDNSRNSCQDSEADEETSP 318
               |  :   :|  : :|||       |     |||| |     : |:  |      :   :   |:|  |   |
Db        244 KTEKRKRSKTSSEESSESEKSDEEEEEKESSPKPKKKPLAVKKLSSDEESEESDVEVLP 303

Qy        319 ----------VFDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENN 368
                        : |  :|::  :: ::  |   :  :  :    :|:                |::
Db        304 QKKKRGAVTLISDSEDEKDQKSESEASDVEEKVSKKKAKKQ----------------ESS 347

Qy        369 AGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISKSE----DRRKEQCEVRH 424
                 ||       :    |: |   | ||:   | :  |       |   ||  :      |   : : |  |
Db        348 ESGSD------SSEGSITVN-RKSKKKEKPEKKKKGIIMDSSKLQKETIDAERAEKE-RR 399

Qy        425 QRTERKIPKYIPPNLPPEKKWLGTPIEE-------------MRKM-------PRCGIHLP 464
              :|  |:|                :|::  |   :||             ||:                 :
Db        400 KRLEKK-----------QKEFNGIVLEEGEDLTEMLTGTSSQRKLKSVVLDPDSSTVDEE 448

Qy        465 SLRPSASHTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGER-- 522
              |  :|    |    ||          :  ||    |                        |  :|     |: :|
Db        449 SKKPVEVHNSLVR--------ILKPHQAH-------------GIQFMYDCACESLDRLD 486

Qy        523 TAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLY 582
              | ||                         || :  :     |         :: |     ||          ::
Db        487 TEGS--------------------GGILAHCMGLGKT---LQVITFLHTVL-------MH 516

Qy        583 QSILPDMVKIALCLP-NICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKM 641
              : |       :: :  : :| |:     |     |:         |    :  |: :: |
Db        517 EKIGEKCKRVLVVVPKNVI---INWFKEFQKWLVD-NDEELDTIDVN------------ 559

Qy        642 KSEYSSYPDINFNR--LFEGRSSRKPEKLKTLFCYFRRVT------EKKPTGLVTFTRQS 693
                 |   ||   |     |    |         ||: |    :      :  || :|            :|||          ::
Db        560 --ELDSYKTIEDRRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKKKKPKNRNRRLEKA 617

Qy        694 LEDF--------PEWERCE-----------------KPLTRLHVTYEGT-IEGN------ 721
                |||           |:      |:                |  ||:   :     ||  :: |
Db        618 KEDFRKYLQNPGPDMVVCDEAHKLKNDDSALSKCMVKILTKRRICLTGTPLQNNLMEYHC 677

Qy        722 -----GRGML--QVDFANRFV 735
                        |:|   :  :||||||
Db        678 MVNFVKPGLLGTKTEFANRFV 698

Search completed: May 26, 2004, 18:41:15
Job time : 19.5091 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:         May 26, 2004, 17:10:29 ; Search time 169.229 Seconds
                                          (without alignments)
                                          13416.302 Million cell updates/sec

Title:          US-09-302-812-1
Perfect score:  7242
Sequence:       1 accggaaagtgaacgaagcc..........aaattttcattaacaaaaaa 4070

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters:      2298626

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150056_5800/app_query.fasta_1.12437
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=6333148_@CGN_1_1_79_@runat_26052004_150056_5800
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_AA:*
                1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5184	71.6	977	9	US-09-973-451-2	Sequence 2, Appli
2	4635.5	64.0	976	9	US-09-973-451-4	Sequence 4, Appli
3	4332.5	59.8	968	9	US-09-973-451-6	Sequence 6, Appli
4	1063.5	14.7	768	9	US-09-973-451-8	Sequence 8, Appli
5	594	8.2	546	12	US-10-425-114-60000	Sequence 60000, A
6	465	6.4	726	9	US-09-973-451-10	Sequence 10, Appl
7	205.5	2.8	200	12	US-10-424-599-184988	Sequence 184988,
8	184	2.5	180	12	US-10-424-599-156445	Sequence 156445,
9	184	2.5	2665	9	US-09-864-761-34248	Sequence 34248, A
10	184	2.5	3664	12	US-10-263-929-143	Sequence 143, App
11	184	2.5	3664	14	US-10-177-293-423	Sequence 423, App
12	175.5	2.4	1163	12	US-10-221-278-197	Sequence 197, App
13	175.5	2.4	1163	15	US-10-291-172-197	Sequence 197, App
14	174.5	2.4	1462	14	US-10-287-218-17	Sequence 17, Appl
15	173.5	2.4	1163	10	US-09-932-257A-3	Sequence 3, Appli
16	173	2.4	1210	14	US-10-205-823-264	Sequence 264, App
17	172.5	2.4	2828	12	US-10-072-012-607	Sequence 607, App
18	170.5	2.4	2828	9	US-09-905-129-21	Sequence 21, Appl
19	170.5	2.4	2828	9	US-09-991-630-21	Sequence 21, Appl
20	170.5	2.4	2828	12	US-10-072-012-608	Sequence 608, App
21	170.5	2.4	2828	12	US-10-454-351-21	Sequence 21, Appl
22	170.5	2.4	2828	12	US-09-802-318-21	Sequence 21, Appl
23	170.5	2.4	2828	14	US-10-176-847-54	Sequence 54, Appl
24	170.5	2.4	2828	14	US-10-177-293-110	Sequence 110, App
25	170.5	2.4	2828	14	US-10-301-822-49	Sequence 49, Appl
26	170.5	2.4	2828	14	US-10-032-189-126	Sequence 126, App
27	170.5	2.4	2828	15	US-10-295-027-58	Sequence 58, Appl
28	170.5	2.4	2828	15	US-10-295-027-1175	Sequence 1175, Ap
29	170.5	2.4	6642	15	US-10-369-493-5013	Sequence 5013, Ap
30	169.5	2.3	913	13	US-10-117-604-4	Sequence 4, Appli
31	167	2.3	31	9	US-09-973-451-11	Sequence 11, Appl
32	165.5	2.3	2515	12	US-10-276-774-2651	Sequence 2651, Ap
33	164.5	2.3	300	12	US-10-424-599-233915	Sequence 233915,
34	164.5	2.3	1804	14	US-10-275-140-2	Sequence 2, Appli
35	164	2.3	1390	12	US-10-092-900A-224	Sequence 224, App
36	163	2.3	747	12	US-10-425-114-67803	Sequence 67803, A
37	162.5	2.2	1109	12	US-10-425-114-72939	Sequence 72939, A
38	162	2.2	748	9	US-09-864-761-43244	Sequence 43244, A
39	162	2.2	1404	9	US-09-811-045A-1	Sequence 1, Appli
40	161.5	2.2	5560	12	US-10-263-929-142	Sequence 142, App
41	161	2.2	691	15	US-10-108-260A-4345	Sequence 4345, Ap
42	161	2.2	951	14	US-10-379-616-8	Sequence 8, Appli
43	161	2.2	1420	14	US-10-379-616-4	Sequence 4, Appli
44	160	2.2	498	12	US-10-424-599-196154	Sequence 196154,
45	160	2.2	748	15	US-10-108-260A-3747	Sequence 3747, Ap

ALIGNMENTS

```
RESULT 1
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
```

```
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-973-451-2

Alignment Scores:
Pred. No.:              0              Length:         977
Score:                  5184.00        Matches:        977
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            71.58%         Indels:         0
DB:                     9              Gaps:           0

US-09-302-812-1 (1-4070) x US-09-973-451-2 (1-977)

Qy     258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy     318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy     378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy     438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy     498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy     558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy     618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy     678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy     738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy     798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy     858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy     918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy     978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy    1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy    1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy    1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy    1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy    1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy    1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy    1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy    1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy    1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy    1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy    1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy    1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy    1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy    1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy    1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy    1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy    2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy    2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy    2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660
```

```
Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGGGCC 3188
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977
```

RESULT 2
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-973-451-4

```
Alignment Scores:
Pred. No.:                 0           Length:          976
Score:                     4635.50     Matches:         874
Percent Similarity:        92.83%      Conservative:    32
Best Local Similarity:     89.55%      Mismatches:      69
Query Match:               64.01%      Indels:          1
DB:                        9           Gaps:            1

US-09-302-812-1 (1-4070) x US-09-973-451-4 (1-976)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              |||:::|||||||||||||||||||||||||||||            ||||||   |||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||   |||||||||||||||||||||||||||||||      |||||||||||||||
Db         21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              |||  |||||||||   ||||||||||||||||||||||||||           |||
Db         40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              |||||||||||||||||||||||||||||||||   ||||||||||||||   ||||||
Db         80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              |||||||||:::||||||||   ||||||||||||||||     ||||||||||
Db        120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||   |||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||   |||:::|||:::||||||||||||||||||        ||||||||||||||
Db        160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              |||||||||||||||||||:::|||:::|||                ||||||||||||||||
Db        180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
              :::||||||||:::||||||||||||||||||||||||||   ||||||||||||||||||
Db        200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
```

```
Qy    918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db    220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy    978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db    240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy   1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db    260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279

Qy   1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db    280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy   1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db    300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy   1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db    320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy   1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db    340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy   1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db    360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy   1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db    380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy   1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db    400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419

Qy   1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db    420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy   1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db    440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459

Qy   1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db    460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy   1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db    480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499

Qy   1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db    500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy   1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db    520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy   1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db    540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy   1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db    560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy   1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db    580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy   2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db    600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy   2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db    620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy   2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db    640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy   2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db    660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy   2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db    680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy   2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db    700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy   2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db    720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy   2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db    740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy   2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db    760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy   2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db    780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

Qy   2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
Db    800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

Qy   2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
Db    820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy   2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
Db    840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy   2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
Db    860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy   2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
Db    880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy   2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
Db    900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy   3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
Db    920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy   3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
```

```
Db        940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
Db        960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975

RESULT 3
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-973-451-6

Alignment Scores:
Pred. No.:               3.82e-304      Length:          968
Score:                   4332.50        Matches:         820
Percent Similarity:      89.86%         Conservative:    57
Best Local Similarity:   84.02%         Mismatches:      90
Query Match:             59.82%         Indels:          9
DB:                      9              Gaps:            5

US-09-302-812-1 (1-4070) x US-09-973-451-6 (1-968)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACGCCGCTGCA 317
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db        217 LysProThrValGlyAspGlyGln---------AlaArgSerAsnCysLysCysSerGly 233

Qy        978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy       1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db        314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332

Qy       1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db        333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy       1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db        353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy       1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db        372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy       1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db        392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy       1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db        412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy       1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db        432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy       1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db        452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy       1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db        472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy       1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
```

```
Db     492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy    1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
           ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db     512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy    1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
           |||||||||||||||||:::  ||||||||||||||||||||||||||||||||||||||||
Db     532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy    1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
           ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db     552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy    1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy    2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy    2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
           ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy    2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy    2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy    2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy    2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
           ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db     692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy    2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
           ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db     712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy    2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
           |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db     732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy    2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy    2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy    2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
           |||||||||||||||||||||||||    ||||||:::||||||||||||||||||   |||
Db     792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy    2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy    2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
           |||:::||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db     832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy    2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy    2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
           ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db     872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy    2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy    3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
           ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||||
Db     912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy    3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy    3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
           |||||||||         :::||||||:::     |||||||||:::     |||
Db     952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

RESULT 4
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-973-451-8

Alignment Scores:
Pred. No.:               6.12e-68        Length:         768
Score:                   1063.50         Matches:        244
Percent Similarity:      57.31%          Conservative:   93
Best Local Similarity:   41.50%          Mismatches:     198
Query Match:             14.69%          Indels:         53
DB:                      9               Gaps:           15

US-09-302-812-1 (1-4070) x US-09-973-451-8 (1-768)

Qy    1506 AGAGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACA 1565
           |||:::   |||:::   |||               :::   |||   :::   :::
Db      45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy    1566 GAAAGGAAGATCCCTAAATACATTCCACCTCACCTT------TCTCCAGATAAGAAATGG 1619
                                    |||   :::|||   |||   |||      |||
Db      63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy    1620 CTTGGAACTCCTATTGAGGAGATG------AGGAGAATGCCAAGGTGTGGGATCCGGCTG 1673
              |||      :::|||   :::         |||   |||            |||
Db      75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy    1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATA 1733
           ||||||:::                     |||   |||      :::          |||   |||
```

```
Db     95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln---------LeuProIle 111
Qy   1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTAAAGATTTGTGGGACAACAAGCATGTT 1793
Db    112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130
Qy   1794 AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCT 1853
Db    131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150
Qy   1854 GCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAG 1913
Db    151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170
Qy   1914 AACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1973
Db    171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190
Qy   1974 GCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAG 2033
Db    191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210
Qy   2034 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCA 2093
Db    211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230
Qy   2094 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCACAGGAACAGATTGCC 2153
Db    231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250
Qy   2154 AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA 2210
Db    251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270
Qy   2211 TCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCA 2270
Db    271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290
Qy   2271 AGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2324
Db    291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310
Qy   2325 AAAAAA------------CCCACTGGGTTGGTGACATTCACAAGACAGAGT------CTT 2366
Db    311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330
Qy   2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTG------ACTCGACTGCATGTC 2420
Db    331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350
Qy   2421 ACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
Db    351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370
Qy   2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
Db    371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390
Qy   2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
Db    391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410
Qy   2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
Db    411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430
Qy   2661 TGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACTGAG 2720
Db    431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450
Qy   2721 ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAG 2780
Db    451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470
Qy   2781 ATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGT----------------- 2822
Db    471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490
Qy   2823 CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCC 2882
Db    491 ProGly-----------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502
Qy   2883 TTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG 2942
Db    503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522
Qy   2943 CGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002
Db    523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542
Qy   3003 CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGA 3062
Db    543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561
Qy   3063 TATTACAATGAAGAATGCAGAAACTGCTCCACCCCCGGACCA----------------- 3104
Db    562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
Qy   3105 GACATCAAGCTTTATCCATTCATATACCATGCAGTTGAGTCCTGTACACAGACCACCAAC 3164
Db    582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598
Qy   3165 CAGCCGGGACAAAGGACGGGGGCC 3188
Db    599 ValProGlyGluGlyAlaSerAla 606
```

RESULT 5
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000

Alignment Scores:			
Pred. No.:	4.61e-34	Length:	546
Score:	594.00	Matches:	153
Percent Similarity:	49.78%	Conservative:	72
Best Local Similarity:	33.85%	Mismatches:	157
Query Match:	8.20%	Indels:	70
DB:	12	Gaps:	12

US-09-302-812-1 (1-4070) x US-10-425-114-60000 (1-546)

```
Qy   1890 CTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCA 1949
Db     79 ValSerArgValAlaSerGlyAspValLeuAlaAspAlaLeuThrAspLeuArgLeuAla 98
Qy   1950 TATTCT-----------------AAGAAATGGGACTTTACAGCTTTGATTGATTTCTGG 1991
```

```
Db     99 LeuAlaLeuProAlaLeuProProArgAlaAlaAspGlyLeuAlaLeu------PhePhe 116

Qy   1992 GATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051

Db    117 AspAspLeuLeuSerArgAlaGlnAlaArgGlyTrpPheSerGluValValProAsnLeu 136

Qy   2052 GTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAG 2111

Db    137 AlaArgLeuLeuLeuArgLeuProThrLeuLeuGluAspHisTyr------------Ala 152

Qy   2112 AAGATGAATCATTCC----------------------------------------ATC 2129

Db    153 LysAlaGlyHisGlyAlaSerGlyLeuArgValMetAlaSerGlnAspAlaGlyValVal 172

Qy   2130 ACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCA 2189

Db    173 LeuLeuSerGlnGluLeuValAlaAlaLeuLeuThrCysAlaLeuPheCysLeuPhePro 192

Qy   2190 CGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGG 2249

Db    193 ThrAlaGlyArgAlaGln------------AlaCysLeuProThrIleAsnPheAspGly 208

Qy   2250 TTGTTT---------------GAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACG 2294

Db    209 LeuPheAlaProLeuIleHisAsnAlaArgGlnSerGlnGlu---GlnLysValArgCys 227

Qy   2295 CTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACA 2354

Db    228 LeuValHisTyrPheGluArgValThrAspSerMetProThrGlyLeuValSerPheGlu 247

Qy   2355 AGACAGAGTCTT---------------------------------GAAGATTTTCCAGAG 2381

Db    248 ArgLysValLeuProArgArgAlaLeuSerAspGlyValProTyrProAspIleHisAla 267

Qy   2382 TGGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTACGAAGGTACCATAGAA 2441

Db    268 TrpValAlaSerSerAlaProLeuCysGlnPheThrValPheSerSerGlyPheIleGlu 287

Qy   2442 GGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAACCGTTTCGTTGGAGGTGGTGTA 2501

Db    288 AspGluGluGlnGluAlaLeuGlnValAspPheAlaAsnLysTyrLeuGlyGlyGlyAla 307

Qy   2502 ACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAACCCTGAGTTGATTGTT 2561

Db    308 LeuSerTrpGlyCysValGlnGluGluIleArgPheMetIleAsnProGluLeuIleVal 327

Qy   2562 TCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTTATCATCACAGGTACTGAG 2621

Db    328 GlyMetLeuPheLeuSerCysMetGluAspAsnGluAlaIleGluIlePheGlyAlaGlu 347

Qy   2622 CAGTACAGTGAATACACAGGCTATGCCGAAACATACCGCTGGGCCCGGAGCCATGAAGAC 2681

Db    348 ArgPheSerGlnTyrMetGlyTyrGlySerSerPheArgPheValGlyAspTyrLeuAsp 367

Qy   2682 AGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTC 2741

Db    368 ThrLysProPheAspSerMetGlyArgArgArgThrArgIleValAlaIleAspAlaLeu 387

Qy   2742 CACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAG 2801

Db    388 AspCysProAlaArgLeu---HisTyrGluSerGlyCysLeuLeuArgGluValAsnLys 406

Qy   2802 GCTTACTGTGGATTTCTTCGT------------------------------------- 2822

Db    407 AlaPheCysGlyPhePheAspGlnSerLysHisHisLeuTyrAlaLysLeuPheGlnAsp 426

Qy   2823 --------------------CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACA 2861

Db    427 LeuHisAsnLysAspAspPheSerSerIleAsnSerSerGluTyrValGlyValSerThr 446

Qy   2862 GGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATC 2921

Db    447 GlyAsnTrpGlyCysGlyAlaPheGlyGlyAsnProGluIleLysSerMetIleGlnTrp 466

Qy   2922 CTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTT---TATTTCACCTTTGGGGACTCAGAA 2978

Db    467 IleAlaAlaSerGlnAlaLeuArgProPheValAsnTyrTyrThrPheGluAspValSer 486

Qy   2979 CTGATGAGAGACATTTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACTGTTGGA 3038

Db    487 LeuGlnArg---LeuGluGluValIleGlnTrpIleArgLeuHisGlyTrpThrValGly 505

Qy   3039 GAAGTATATAAGCTGCTGCTACGATATTACAATGAA 3074

Db    506 GluLeuTrpHisMetLeuMetGluTyrSerSerGln 517
```

RESULT 6
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-973-451-10

Alignment Scores:

Pred. No.:	1.07e-24	Length:	726
Score:	465.00	Matches:	204
Percent Similarity:	38.19%	Conservative:	116
Best Local Similarity:	24.34%	Mismatches:	282
Query Match:	6.42%	Indels:	236
DB:	9	Gaps:	35

US-09-302-812-1 (1-4070) x US-09-973-451-10 (1-726)

```
Qy    852 GACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCG 911

Db     10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28

Qy    912 AATGCAAAGCAGACGATGGAA------------GATGAACAGGGCAGAGAAGCCAGAAGC 959

Db     29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48

Qy    960 CACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAG 1019

Db     49 LysGlu------------------AspProGluGlu------------------ProLys 56

Qy   1020 GAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGAT--------- 1070

Db     57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76

Qy   1071 --------------------GTTGGTACTGGACTGAAAAATGCC--------------- 1094

Db     77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96

Qy   1095 -----------------------AACAGATTGAATAGACAAGAAAGTAGTCTAGGAAAT 1130
```

```
Db      97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116
Qy    1131 ---------TCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTA--- 1178
Db     117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136
Qy    1179 --------------------------------GATAATTCCAAAAATAGTTGTCAGGAT 1205
Db     137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156
Qy    1206 TCAGAAGCAGATGAAGAGACAAGTCCAGGTTTT------GATGAACAGGAA--------- 1250
Db     157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176
Qy    1251 ------------------------------------------GATAGCAGTTCTGCTCAA 1268
Db     177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196
Qy    1269 ACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGG 1328
Db     197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer--------------- 211
Qy    1329 TCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGA 1388
Db     212 ------------------------------------ThrPheValGlyGluAspSerLys 219
Qy    1389 GCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448
Db     220 Ala------------ThrLysThrValArgThrSerSerSerSerPheLeuSerThr--- 234
Qy    1449 TGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGA 1508
Db     234 ------------------------------------------------------------ 234
Qy    1509 GTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAA 1568
Db     235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254
Qy    1569 AGGAAGATCCCTAAATACATTCCACCTCACCTTTCT------CCAGATAAGAAATGGCTT 1622
Db     255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp------------ 270
Qy    1623 GGAACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTG 1682
Db     271 ------------------------------------------------LeuAsnLysVal 274
Qy    1683 AGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTT 1742
Db     275 AspProAspArgAsnTyr---------------------------ArgTyrCysThrIle 285
Qy    1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTTGTGGGACAACAAGCAT--------- 1790
Db     286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304
Qy    1791 GTTAAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1850
Db     305 IleValLeuProGlnArgTrpArgGluPhe------------AspSerArgGlyArgArg 320
Qy    1851 GCTGCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCC 1910
Db     320 ------------------------------------------------------------ 320
Qy    1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT---------GTGGCATATTCTAAGAAA 1961
Db     321 ---------ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337
Qy    1962 TGGGACTTTACAGCT---------------------TTGATTGATTTCTGGGAT------ 1994
Db     338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAsp 357
Qy    1995 ---AAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051
Db     358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377
Qy    2052 GTG--------------------------------------AAAATTGCACTCTGTCTG 2072
Db     378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397
Qy    2073 CCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACA 2132
Db     398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416
Qy    2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192
Db     417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe--------------- 431
Qy    2193 CGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTG 2252
Db     432 ------AlaArgProAspSerProPheSer------------------PheCysArgIle 443
Qy    2253 TTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGA 2312
Db     444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463
Qy    2313 AGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTC------ACAAGACAGAGTCTT 2366
Db     464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483
Qy    2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTAC 2426
Db     484 AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503
Qy    2427 ------GAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
Db     504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522
Qy    2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
Db     523 GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542
Qy    2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
Db     543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562
Qy    2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
Db     563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582
Qy    2661 TGGGCC---------------CGGAGCCATGAAGACAGGAGCGAA---AGGGACGACTGG 2702
Db     583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602
Qy    2703 CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC--------- 2753
Db     603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622
Qy    2754 ------TACCTCGACCAGTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAGGCTTAC 2807
Db     623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642
Qy    2808 TGTGGATTTCTTCGTCCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAAC 2867
Db     643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp 661
Qy    2868 TGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCA 2927
Db     662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681
Qy    2928 GCTGCTGTAGCTGAGCGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGA 2987
Db     682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701
Qy    2988 GACATTTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAA 3041
Db     702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719
```

```
RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(200)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

Alignment Scores:
Pred. No.:              3.87e-06     Length:        200
Score:                  205.50       Matches:       52
Percent Similarity:     44.07%       Conservative:  26
Best Local Similarity:  29.38%       Mismatches:    70
Query Match:            2.84%        Indels:        29
DB:                     12           Gaps:          4

US-09-302-812-1 (1-4070) x US-10-424-599-184988 (1-200)

Qy       2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192
               ::::::||||||       |||         :::         ||||||   ||||||
Db         33 LeuThrGlnGluThrSerAlaProProProSerArgThrLeuPheCysLeuPhePro--- 51

Qy       2193 CGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCA--------------------- 2231
                                          |||      |||
Db         52 -----------------------ValSerAspArgProGlyIleHisLeuProMetIle 63

Qy       2232 ------GATATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAG 2285
                    |||::::::          |||:::|||       |||    :::
Db         64 AsnSerAspValSerSerGlySerLeuTyrGluAspTyrSerGlnLysProGluSerThr 83

Qy       2286 CTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTG 2345
                       :::      |||    :::|||::::::   :::    |||    |||:::|||
Db         84 ThrGlyArgIleAlaHisTyrGlyGlnArgIleSerSerGluMetProLysGlyIleVal 103

Qy       2346 ACATTCACAAGACAGAGTCTTGAA-----------------------GATTTCCAGAG 2381
              :::|||   |||:::   |||                          :::|||:::
Db        104 SerPheGluArgLysValLeuProPheLysAsnAspSerIleHisIleSerTyrProAsp 123

Qy       2382 ---------TGGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTACGAAGGT 2432
                       |||               |||   |||         |||         |||
Db        124 AlaAsnPheTrpSerThrSerAlaIleProLeuCysArgPheGluValHisSerSerGly 143

Qy       2433 ACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAACCGTTTCGTTGGA 2492
                 ||||||            |||      :::||||||   ||||||:::::::::|||
Db        144 LeuIleGluAspGlnSerSerGlyAlaAlaGluValAspLeuAlaAsnLysTyrLeuGly 163

Qy       2493 GGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAACCCTGAG 2552
              ||||||            |||   ||||||||||||||||||:::::::::||||||
Db        164 GlyGlyAlaLeuGlyArgGlyCysValGlnGluGluIleArgPheMetValSerProGlu 183

Qy       2553 TTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTT 2603
              |||            ||||||       :::       ||||||   :::
Db        184 LeuGluAlaGlyMetLeuPheLeuProAlaMetAlaAspAsnGluAlaIle 200
```

```
RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

Alignment Scores:
Pred. No.:              0.000133     Length:        180
Score:                  184.00       Matches:       48
Percent Similarity:     44.17%       Conservative:  24
Best Local Similarity:  29.45%       Mismatches:    41
Query Match:            2.54%        Indels:        50
DB:                     12           Gaps:          4

US-09-302-812-1 (1-4070) x US-10-424-599-156445 (1-180)

Qy       2790 GAGCTTAACAAGGCTTACTGTGGATTTCTT---------------CGTCCT--------- 2825
              |||:::|||||||||:::||||||||||||                  :::|||
Db          7 GluIleAsnLysAlaPheCysGlyPheLeuTyrGlnCysLysTyrGlnProTyrGlnLys 26

Qy       2825 ------------------------------------------------------------ 2825

Db         27 IleLeuGlnGluAsnGlyCysThrSerAlaLeuPheTyrAlaAlaThrSerThrSerMet 46

Qy       2825 ------------------------------------------------------------ 2825

Db         47 GluThrAspGluGlyGluIleSerAsnHisLysIleThrAsnSerGlnAsnAspTyrHis 66

Qy       2826 GGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTT 2885
              |||:::         |||         ||||||||||||||||||||||||||||||
Db         67 GlyMetAspGlnGlyAsnAsnIleGlyValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 86

Qy       2886 GGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGA 2945
              |||||||||         :::|||   :::||||||   |||||||||:::   |||   |||
Db         87 GlyGlyAspProGluValLysThrIleIleGlnTrpLeuAlaAlaSerGlnAlaLeuArg 106

Qy       2946 GAC---GTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002
                 :::   |||:::|||||||||         |||   ::::::   :::      :::
Db        107 ProPheIleAlaTyrTyrThrPheGly---LeuGluAlaLeuGlnSerLeuAspGluVal 125

Qy       3003 CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGA 3062
                 ::::::   ::::::   |||||||||:::::::::   :::|||:::
Db        126 AlaHisTrpIleLeuSerGlnArgTrpThrValGlyAspLeuTrpAsnMetLeuIleGlu 145

Qy       3063 TATTACAATGAAGAATGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCA 3122
              |||            :::      :::      |||               |||   |||
Db        146 TyrSerIleAsnArgSerLysGlyGluThrAsnValGlyPheLeuGlnTrpLeuLeuPro 165

Qy       3123 TTCATATAC 3131
                 ||||||
Db        166 SerIleTyr 168

RESULT 9
US-09-864-761-34248
; Sequence 34248, Application US/09864761
```

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
;   LENGTH: 2665
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034555.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  12
;   OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 3.00e-10
US-09-864-761-34248

Alignment Scores:
Pred. No.:              0.000342      Length:         2665
Score:                  184.00        Matches:        126
Percent Similarity:     33.69%        Conservative:   92
```

```
Best Local Similarity:  19.47%      Mismatches:    268
Query Match:            2.54%       Indels:        161
DB:                     9           Gaps:          24

US-09-302-812-1 (1-4070) x US-09-864-761-34248 (1-2665)

Qy        6 AAAGTGAACGAAGCCCGAATCAGAACGGCTCATCCTGAGGCTGGT--------------- 50
              :::               :::|||:::   :::         |||
Db     1281 ArgThrArgArgArgAsnValArgSerValTyrAlaThrMetGlyAspHisGluAsnArg 1300

Qy       51 ---------AGGGTGCCGGTGGAAGAGGGAAGGCAGGCGTCTGGATAGGGCCTGGTTCGG 101
                  :::    |||||||||:::    |||                  :::   |||
Db     1301 SerProValLysGluProValGluGlnProArg-----------------ValThrArg 1314

Qy      102 GAGGCTGTCAGAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGA 161
            :::   :::  |||:::|||||||||:::    :::|||
Db     1315 LysArgLeu-GluArgGluLeuGlnGluAlaAlaAlaValPro---------------Th 1329

Qy      162 GGCACCGAGGAGGGGGCGCAGTCCGTCCCTCCCAGGGTTAGTG---------------AA 206
                 ||||||||||||||     |||                              ||
Db     1329 rThrProArgArgGlyArgProProLysThrArgArgArgAlaAspGluGluGluGluAs 1349

Qy      207 TGAGGCTCTACGCCCG-----------------------GGCTGGCCCGGAGACTCAGT 242
            ||||||       |||                        ||||||   |||
Db     1349 nGluAlaLysGluProAlaGluThrLeuLysProProGluGlyTrp--ArgSerProArg 1368

Qy      243 GCTGCGGGTCCCAGCATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC 302
            :::             :::   ||||||                        |||:::
Db     1369 SerGlnLysThrAlaAlaGlyGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 1388

Qy      303 TGGGACGCCGCTGCAACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGG 362
               |||      ||||||   |||   |||   :::::                |||
Db     1389 ValAsp------AlaThrArgProGluAlaThrThrGluVal-----------GlyPro 1402

Qy      363 CAGAGGCGCGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCG 422
            |||      |||   :::|||               |||          :::
Db     1403 GlnIleGlyValLysGluSerSerMetGluPro---------LysAlaAlaGluGluGlu 1419

Qy      423 TCAGGCTGCGCCCTGGGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTC 482
            :::|||             |||      :::   |||
Db     1420 AlaGlySerGluGlnLysArgAspArgLysAspAlaGly-------------------- 1432

Qy      483 AAACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAA 542
                       |||      :::     :::|||    :::::   ||||||    :::
Db     1433 ---------ThrAspLysAsnProProGluThrAlaProValGluValValGluLysLys 1449

Qy      543 AGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAA 602
                     |||   :::   :::     |||      |||      |||    :::
Db     1450 ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 1469

Qy      603 AAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTT 662
               |||                              |||:::||||||
Db     1470 ValAsp---------------LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 1484

Qy      663 GATAAGTCACCAGTTGAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATG 722
              :::            :::      :::      :::::            |||:::
Db     1485 ProArgGlyAlaAlaAlaGlnAlaGlyGlu------ArgGluSerGlyValValAlaVal 1502

Qy      723 TGTAAGTGGCAGAATGAAGGGCCACACTCAGAA-----------CGGCTTTTGGAAAGT 770
                  ::::::|||   |||        |||                |||:::|||
Db     1503 SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 1522

Qy      771 GAACCTCCAGCGGTAACTCTGGTACCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCG 830
            :::|||                  ||||||:::               :::|||    |||
Db     1523 AspProValAspProAspLysGluProGluLys---------GluAspValSerAlaSer 1539

Qy      831 TCCCCAAAGGATGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTT 890
               |||       :::         |||         :::      :::    :::|||
Db     1540 GlyProSerProGlu------AlaThrGlnLeuAlaLysGlnMetGluLeuGluGlnAla 1557

Qy      891 TTGACACATGTA---AAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGA 947
            :::   |||:::   |||||||||   |||
Db     1558 ValGluHisIleAlaLysLeuAlaGluAlaSer-------------------------- 1568
```

```
Qy      948 GAAGCCAGAAGCCACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGG 1007
                                                |||  :::  |||:::|||      |||
Db     1569 ------------------------AlaSerAlaAlaTyrLysAlaAspAlaProGluGly 1580

Qy     1008 TGTCAGCAGGAGGAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAG 1067
                     |||:::   |||                    |||
Db     1581 LeuAlaProGluAspArgAsp------------LysProAlaHisGlnAlaSerGluThr 1596

Qy     1068 GATGTTGGTACTGGACTGAAAAAT---GCCAACAGATTGAATAGACAAGAAAGTAGTCTA 1124
            ::::::      :::   :::      |||  ::::::  :::        :::
Db     1597 GluLeuAlaAlaAlaIleGlySerIleIleAsnAspIleSerGlyGluProGluAsnPhe 1616

Qy     1125 GGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA--------------------------- 1157
                     ||||||:::        ||||||:::
Db     1617 ProAlaProProProTyrProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla 1636

Qy     1158 ---------------CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAG 1202
                              |||       |||:::   |||   :::   :::
Db     1637 GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu 1656

Qy     1203 GATTCAGAAGCAGATGAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCT 1262
            ::::::||||||         |||:::|||                |||      |||:::
Db     1657 GluThrGluAlaAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla 1676

Qy     1263 GCTCAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGG 1322
                  |||       |||   :::|||            |||            :::
Db     1677 GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro 1696

Qy     1323 AAGCGGTCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGA------ 1376
            :::      :::|||      |||:::   |||      :::   :::|||
Db     1697 GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln 1716

Qy     1376 ------------------------------------------------------------ 1376

Db     1717 LysThrThrArgSerArgArgLysArgAsnThrAsnLysLysValValAlaProValGlu 1736

Qy     1377 -----------------------------GGAGAGAGTCGAGCTGGAATGAATGATGTG 1406
                                         |||||||||   |||
Db     1737 SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr 1756

Qy     1407 AATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCTAAGCAA 1466
                 :::   |||   :::                     |||   ::::::   |||
Db     1757 ThrValGlnHisProGluAlaProGlnGlu---------GluLysGlnSerGluLysPro 1773

Qy     1467 CATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAAGCAGAGGAC 1526
            |||                     |||            ::::::|||      |||:::
Db     1774 HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn 1793

Qy     1527 AAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATCCCTAAATAC 1586
                  :::|||            :::         |||:::      :::|||
Db     1794 SerSerGlnGluIleSerValGluGluArgThrProThrLysAlaSerValProProAsp 1813

Qy     1587 ATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGG 1646
            :::||||||         |||                   |||::::::|||
Db     1814 LeuProProProProGlnPro------------------AlaProValAspGlu------ 1825

Qy     1647 AGAATGCCAAGGTGTGGGATCCGGCTG------------------------------CCT 1676
                  |||:::         |||:::                              |||
Db     1826 ---GluProGlnAlaArgPheArgValHisSerIleIleGluSerAspProValThrPro 1844

Qy     1677 CCCTTGAGACCATCT 1691
            |||      ||||||
Db     1845 ProSerAspProSer 1849
```

RESULT 10
US-10-263-929-143
; Sequence 143, Application US/10263929
; Publication No. US20040067535A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Jaeseob
; APPLICANT: Galant, Ron
; TITLE OF INVENTION: Alzheimer's Disease Linked Genes
; FILE REFERENCE: LSD-07417
; CURRENT APPLICATION NUMBER: US/10/263,929
; CURRENT FILING DATE: 2002-10-03
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 143
; LENGTH: 3664
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-263-929-143

Alignment Scores:			
Pred. No.:	0.000382	Length:	3664
Score:	184.00	Matches:	126
Percent Similarity:	33.69%	Conservative:	92
Best Local Similarity:	19.47%	Mismatches:	268
Query Match:	2.54%	Indels:	161
DB:	12	Gaps:	24

US-09-302-812-1 (1-4070) x US-10-263-929-143 (1-3664)

```
Qy        6 AAAGTGAACGAAGCCCGAATCAGAACGGCTCATCCTGAGGCTGGT--------------- 50
            :::               :::|||:::   :::               |||
Db     1898 ArgThrArgArgArgAsnValArgSerValTyrAlaThrMetGlyAspHisGluAsnArg 1917

Qy       51 ---------AGGGTGCCGGTGGAAGAGGGAAGGCAGGCGTCTGGATAGGGCCTGGTTCGG 101
                     :::   |||||||||:::   |||                  :::   |||
Db     1918 SerProValLysGluProValGluGlnProArg------------------ValThrArg 1931

Qy      102 GAGGCTGTCAGAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGA 161
            :::   :::   |||:::|||||||||:::   :::|||
Db     1932 LysArgLeu-GluArgGluLeuGlnGluAlaAlaAlaValPro---------------Th 1946

Qy      162 GGCACCGAGGAGGGGGCGCAGTCCGTCCCTCCCAGGGTTAGTG---------------AA 206
                |||||||||||||||   |||                                 ||
Db     1946 rThrProArgArgGlyArgProProLysThrArgArgArgAlaAspGluGluGluGluAs 1966

Qy      207 TGAGGCTCTACGCCCG-----------------------GGCTGGCCCGGAGACTCAGT 242
            ||||||      |||                        ||||||   |||
Db     1966 nGluAlaLysGluProAlaGluThrLeuLysProProGluGlyTrp--ArgSerProArg 1985

Qy      243 GCTGCGGGTCCCAGCATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC 302
            :::      :::            ||||||      :::                |||:::
Db     1986 SerGlnLysThrAlaAlaGlyGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 2005

Qy      303 TGGGACGCCGCTGCAACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGG 362
               |||      ||||||   |||   |||   ::::::                  |||
Db     2006 ValAsp------AlaThrArgProGluAlaThrThrGluVal------------GlyPro 2019

Qy      363 CAGAGGCGCGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCG 422
            |||   |||   :::|||         |||                  :::
Db     2020 GlnIleGlyValLysGluSerSerMetGluPro---------LysAlaAlaGluGluGlu 2036

Qy      423 TCAGGCTGCGCCCTGGCCGGGCGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTC 482
            :::|||         |||      :::         |||
Db     2037 AlaGlySerGluGlnLysArgAspArgLysAspAlaGly--------------------- 2049

Qy      483 AAACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAA 542
                        |||      :::      :::|||      ::::::   ||||||   :::
Db     2050 ---------ThrAspLysAsnProProGluThrAlaProValGluValValGluLysLys 2066

Qy      543 AGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAA 602
                     |||   :::   :::   |||      |||      |||   :::
Db     2067 ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 2086

Qy      603 AAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTT 662
               |||                              |||:::||||||
Db     2087 ValAsp---------------LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 2101

Qy      663 GATAAGTCACCAGTTGAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATG 722
               :::         :::      :::         ::::::               |||:::
Db     2102 ProArgGlyAlaAlaAlaGlnAlaGlyGlu------ArgGluSerGlyValValAlaVal 2119
```

```
Qy        723 TGTAAGTGGCAGAATGAAGGGCCACACTCAGAA-----------CGGCTTTTGGAAAGT 770
Db       2120 SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 2139

Qy        771 GAACCTCCAGCGGTAACTCTGGTACCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCG 830
Db       2140 AspProValAspProAspLysGluProGluLys---------GluAspValSerAlaSer 2156

Qy        831 TCCCCAAAGGATGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTT 890
Db       2157 GlyProSerProGlu------AlaThrGlnLeuAlaLysGlnMetGluLeuGluGlnAla 2174

Qy        891 TTGACACATGTA---AAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGA 947
Db       2175 ValGluHisIleAlaLysLeuAlaGluAlaSer-------------------------- 2185

Qy        948 GAAGCCAGAAGCCACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGG 1007
Db       2186 -----------------------AlaSerAlaAlaTyrLysAlaAspAlaProGluGly 2197

Qy       1008 TGTCAGCAGGAGGAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAG 1067
Db       2198 LeuAlaProGluAspArgAsp------------LysProAlaHisGlnAlaSerGluThr 2213

Qy       1068 GATGTTGGTACTGGACTGAAAAAT---GCCAACAGATTGAATAGACAAGAAAGTAGTCTA 1124
Db       2214 GluLeuAlaAlaAlaIleGlySerIleIleAsnAspIleSerGlyGluProGluAsnPhe 2233

Qy       1125 GGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA-------------------------- 1157
Db       2234 ProAlaProProProTyrProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla 2253

Qy       1158 ---------------CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAG 1202
Db       2254 GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu 2273

Qy       1203 GATTCAGAAGCAGATGAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCT 1262
Db       2274 GluThrGluAlaAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla 2293

Qy       1263 GCTCAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGG 1322
Db       2294 GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro 2313

Qy       1323 AAGCGGTCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGA------ 1376
Db       2314 GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln 2333

Qy       1376 ------------------------------------------------------------ 1376
Db       2334 LysThrThrArgSerArgArgLysArgAsnThrAsnLysLysValValAlaProValGlu 2353

Qy       1377 ------------------------------GGAGAGAGTCGAGCTGGAATGAATGATGTG 1406
Db       2354 SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr 2373

Qy       1407 AATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCTAAGCAA 1466
Db       2374 ThrValGlnHisProGluAlaProGlnGlu---------GluLysGlnSerGluLysPro 2390

Qy       1467 CATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAAGCAGAGGAC 1526
Db       2391 HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn 2410

Qy       1527 AAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATCCCTAAATAC 1586
Db       2411 SerSerGlnGluIleSerValGluGluArgThrProThrLysAlaSerValProProAsp 2430

Qy       1587 ATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGG 1646
Db       2431 LeuProProProProGlnPro------------------AlaProValAspGlu------ 2442

Qy       1647 AGAATGCCAAGGTGTGGGATCCGGCTG-----------------------------CCT 1676
Db       2443 ---GluProGlnAlaArgPheArgValHisSerIleIleGluSerAspProValThrPro 2461

Qy       1677 CCCTTGAGACCATCT 1691
Db       2462 ProSerAspProSer 2466
```

RESULT 11
US-10-177-293-423
; Sequence 423, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 423
;   LENGTH: 3664
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-177-293-423

Alignment Scores:			
Pred. No.:	0.000382	Length:	3664
Score:	184.00	Matches:	126
Percent Similarity:	33.69%	Conservative:	92
Best Local Similarity:	19.47%	Mismatches:	268
Query Match:	2.54%	Indels:	161
DB:	14	Gaps:	24

US-09-302-812-1 (1-4070) x US-10-177-293-423 (1-3664)

```
Qy          6 AAAGTGAACGAAGCCCGAATCAGAACGGCTCATCCTGAGGCTGGT--------------- 50
Db       1898 ArgThrArgArgArgAsnValArgSerValTyrAlaThrMetGlyAspHisGluAsnArg 1917

Qy         51 ---------AGGGTGCCGGTGGAAGAGGGAAGGCAGGCGTCTGGATAGGGCCTGGTTCGG 101
Db       1918 SerProValLysGluProValGluGlnProArg-----------------ValThrArg 1931
```

```
Qy        102 GAGGCTGTCAGAGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGA 161
              :::   ::: |||:::|||||||||:::   :::|||
Db       1932 LysArgLeu-GluArgGluLeuGlnGluAlaAlaAlaValPro---------------Th 1946

Qy        162 GGCACCGAGGAGGGGGCGCAGTCCGTCCCTCCCAGGGTTAGTG---------------AA 206
                 |||||||||||||||    |||                               ||
Db       1946 rThrProArgArgGlyArgProProLysThrArgArgAlaAspGluGluGluGluAs 1966

Qy        207 TGAGGCTCTACGCCCG-----------------------GGCTGGCCCGGAGACTCAGT 242
              ||||||     |||                         ||||||  |||
Db       1966 nGluAlaLysGluProAlaGluThrLeuLysProProGluGlyTrp--ArgSerProArg 1985

Qy        243 GCTGCGGGTCCCAGCATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGC 302
              :::         :::        ||||||     :::              |||:::
Db       1986 SerGlnLysThrAlaAlaGlyGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 2005

Qy        303 TGGGACGCCGCTGCAACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGG 362
                 |||      ||||||   |||   |||   ::::::                |||
Db       2006 ValAsp------AlaThrArgProGluAlaThrThrGluVal------------GlyPro 2019

Qy        363 CAGAGGCGCGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCG 422
              |||      |||   :::|||       |||          :::
Db       2020 GlnIleGlyValLysGluSerSerMetGluPro---------LysAlaAlaGluGluGlu 2036

Qy        423 TCAGGCTGCGCCCTGGGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTC 482
              :::|||            |||     :::       |||
Db       2037 AlaGlySerGluGlnLysArgAspArgLysAspAlaGly--------------------- 2049

Qy        483 AAACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAA 542
                       |||     :::    :::|||     ::::::  ||||||   :::
Db       2050 ---------ThrAspLysAsnProProGluThrAlaProValGluValValGluLysLys 2066

Qy        543 AGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAA 602
                          |||   :::   :::   |||      |||     |||   :::
Db       2067 ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 2086

Qy        603 AAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTT 662
                 |||                        |||:::||||||
Db       2087 ValAsp--------------LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 2101

Qy        663 GATAAGTCACCAGTTGAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATG 722
                 :::        :::     :::     ::::::               |||:::
Db       2102 ProArgGlyAlaAlaAlaGlnAlaGlyGlu------ArgGluSerGlyValValAlaVal 2119

Qy        723 TGTAAGTGGCAGAATGAAGGGCCACACTCAGAA------------CGGCTTTTGGAAAGT 770
                    ::::::|||   |||      |||                |||:::|||
Db       2120 SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 2139

Qy        771 GAACCTCCAGCGGTAACTCTGGTACCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCG 830
              :::|||                  ||||||:::             :::|||      |||
Db       2140 AspProValAspProAspLysGluProGluLys---------GluAspValSerAlaSer 2156

Qy        831 TCCCCAAAGGATGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTT 890
                 |||      :::      |||         :::   :::   :::|||
Db       2157 GlyProSerProGlu------AlaThrGlnLeuAlaLysGlnMetGluLeuGluGlnAla 2174

Qy        891 TTGACACATGTA---AAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGA 947
              :::   |||:::   |||||||||   |||
Db       2175 ValGluHisIleAlaLysLeuAlaGluAlaSer--------------------------- 2185

Qy        948 GAAGCCAGAAGCCACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGG 1007
                                       |||   :::   |||:::|||         |||
Db       2186 -----------------------AlaSerAlaAlaTyrLysAlaAspAlaProGluGly 2197

Qy       1008 TGTCAGCAGGAGGAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAG 1067
                 |||:::   |||                  |||
Db       2198 LeuAlaProGluAspArgAsp------------LysProAlaHisGlnAlaSerGluThr 2213

Qy       1068 GATGTTGGTACTGGACTGAAAAAT---GCCAACAGATTGAATAGACAAGAAAGTAGTCTA 1124
              ::::::         :::   :::      |||   ::::::   :::      :::
Db       2214 GluLeuAlaAlaAlaIleGlySerIleIleAsnAspIleSerGlyGluProGluAsnPhe 2233

Qy       1125 GGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA--------------------------- 1157
                                                    ||||||:::     ||||||:::
Db       2234 ProAlaProProProTyrProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla 2253

Qy       1158 ---------------CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAG 1202
                             |||      |||:::   |||   :::   :::
Db       2254 GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu 2273

Qy       1203 GATTCAGAAGCAGATGAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCT 1262
              ::::::||||||      |||:::|||                      |||   |||:::
Db       2274 GluThrGluAlaAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla 2293

Qy       1263 GCTCAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGG 1322
                 |||   |||   :::|||               |||            :::
Db       2294 GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro 2313

Qy       1323 AAGCGGTCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGA------ 1376
              :::      :::|||   |||:::   |||         :::   :::|||
Db       2314 GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln 2333

Qy       1376 ------------------------------------------------------------ 1376

Db       2334 LysThrThrArgSerArgArgLysArgAsnThrAsnLysLysValValAlaProValGlu 2353

Qy       1377 -----------------------------GGAGAGAGTCGAGCTGGAATGAATGATGTG 1406
                                          ||||||||   |||
Db       2354 SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr 2373

Qy       1407 AATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCTAAGCAA 1466
                 :::   |||   :::                  |||   ::::::   |||
Db       2374 ThrValGlnHisProGluAlaProGlnGlu---------GluLysGlnSerGluLysPro 2390

Qy       1467 CATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAAGCAGAGGAC 1526
              |||                        |||         ::::::|||   |||:::
Db       2391 HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn 2410

Qy       1527 AAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATCCCTAAATAC 1586
                 :::|||      :::      |||:::      :::|||
Db       2411 SerSerGlnGluIleSerValGluGluArgThrProThrLysAlaSerValProProAsp 2430

Qy       1587 ATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGG 1646
              :::||||||      |||                  |||::::::|||
Db       2431 LeuProProProProGlnPro------------------AlaProValAspGlu------ 2442

Qy       1647 AGAATGCCAAGGTGTGGGATCCGGCTG----------------------------CCT 1676
                 |||:::   |||:::                                    |||
Db       2443 ---GluProGlnAlaArgPheArgValHisSerIleIleGluSerAspProValThrPro 2461

Qy       1677 CCCTTGAGACCATCT 1691
              |||   ||||||
Db       2462 ProSerAspProSer 2466
```

RESULT 12
US-10-221-278-197
; Sequence 197, Application US/10221278
; Publication No. US20040034208A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/221,278
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07

```
;  NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
;   LENGTH: 1163
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-221-278-197

Alignment Scores:
Pred. No.:              0.00105      Length:        1163
Score:                  175.50       Matches:       152
Percent Similarity:     33.85%       Conservative:  108
Best Local Similarity:  19.79%       Mismatches:    256
Query Match:            2.42%        Indels:        252
DB:                     12           Gaps:          34

US-09-302-812-1 (1-4070) x US-10-221-278-197 (1-1163)

Qy     59 GGTGGAAGAGGGAAGGCAGGCGTCTGG------ATAGGGCCTGGTTCGGGAGGCTGTCAG 112
Db    100 GlyGlySerHisGlnSerSerLysTrpThrProValGlyProAlaProSerThrSerGln 119

Qy    113 AGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGAGGCACCGAGGA 172
Db    120 SerGlnLys---ArgSerSerGlyLeuGlnSerGlyHisSerSerGlnArg-ThrSerAl 138

Qy    173 GGGGGCG----------------------------------------------------- 179
Db    138 aGlySerSerSerGlyThrAsnSerSerGlyGlnArgHisAspArgGluSerTyrAsnAs 158

Qy    180 ---------------------------CAGTCCGTCCCTCCCAGGGTTAGTGAATGAGG 211
Db    158 nSerGlySerSerSerArgLysLysGlyGlnHisGlySerGluHisSerLysSerArgSe 178

Qy    212 CTCTACGCCCGGGCTGGCCCGGAGACTCAGTGCTGCGGGTCCCAGCATGAGTGCGGGCCC 271
Db    178 rSerSerProGlyLysProGlnAlaValSerSerLeuAsnSerSerHisSerArgSerHi 198

Qy    272 CGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCAACTTCTCCGCCGGC 331
Db    198 sGlyAsnAspHisHisSerLysGluHisGlnArgSerLysSerProArgAspProAspAl 218

Qy    332 CGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTCGATTCCAAGGACGC 391
Db    218 aAsnTrpAspSerProSer----------------------------------------- 224

Qy    392 TCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTGGGCCGGGCGGGACA 451
Db    225 -------------ArgValProPheSerSer----------------------GlyGl 232

Qy    452 GCAC--------------------------CGAGGCAGCGCCACCTCTCTTGTTTTCAA 484
Db    232 nHisSerThrGlnSerPheProProSerLeuMetSerLysSerAsnSerMetLeuGlnLy 252

Qy    485 ACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAA-------- 536
Db    252 sProThrAlaTyrValArgProMetAsp------GlyGlnGluSerMetGluProLysLe 270

Qy    537 ----TCAGAAAGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAG 592
Db    270 uSerSerGluHisTyrSerSerGlnSerHisGlyAsnSerMetThrGluLeuLysProSe 290

Qy    593 TTCTGTACAAAAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCA 652
Db    290 rSer--------------------LysAlaHisLeuThrLysLeuLysIleProSerGl 303

Qy    653 G---------------CTAGGTTTTGATAAGTCACCAGTTGAAAAAGGTACACAGTATTT 697
Db    303 nProLeuAspAlaSerAlaSerGlyAspValSerCysValAsp---------GluIleLe 320

Qy    698 GAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAATGAAGGGCCACACTCAGAACG 757
Db    320 uLysGlu----------------------------------------------------- 322

Qy    758 GCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTA----------------------- 794
Db    323 -MetThrHisSerTrpProProProLeuThrAlaIleHisThrProCysLysThrGluPr 342

Qy    794 ------------------------------------------------------------ 794
Db    342 oSerLysPheProPheProThrLysGluSerGlnGlnSerAsnPheGlyThrGlyGluGl 362

Qy    795 -------------CCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGA 841
Db    362 nLysArgTyrAsnProSerLysThrSerAsnGlyHisGlnSerLysSerMetLeuLysAs 382

Qy    842 TGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTT---TTGACACA 898
Db    382 pAspLeuLysLeuSerSerSerGluAspSerAspGlyGluGlnAspCysAspLysThrMe 402

Qy    899 TGTAAAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAG 958
Db    402 tProArgSerThrProGlySerAsnSerGluProSerHisHisAsnSerGluGlyAlaAs 422

Qy    959 CCACCAGAAGTGTGGCAAGGCTTGCCATCCT---GCAGAAGCCTGTGCAGGGTGTCAGCA 1015
Db    422 pAsnSerArgAspAspSerSerSerHisSerGlySerGluSerSerSerGlySerAspSe 442

Qy   1016 GGAGGAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGG 1075
Db    442 rGlu---------------SerGluSerSerSerSerAspSerGluAlaAsnGluProSe 457

Qy   1076 TACTGGA-----------------------------------------CTGAAAAAT-- 1091
Db    457 rGlnSerAlaSerProGluProGluProProProThrAsnLysTrpGlnLeuAspAsnTr 477

Qy   1092 -GCCAACAGATTGAATAGACAA-----------------------GAAAGTAGTCTAGG 1126
Db    477 pLeuAsnLysValAsnProHisLysValSerProAlaSerSerValAspSerAsnIlePr 497

Qy   1127 AAATTCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTC 1186
Db    497 oSerSerGlnGlyTyrLysLysGluGlyArgGluGln-----------------GlyTh 511

Qy   1187 CAAAAATAGTTGTCAGGATTCAGAAGCAGATGAAGAGACAAGT---------CCAGGTTT 1237
Db    511 rGlyAsnSerTyrThrAspThrSerGlyProLysGluThrSerSerAlaThrProGlyAr 531

Qy   1238 TGATGAA---------CAGGAAGATAGCAGTTCTGCTCAAACAGCAAATAAACCTTCAAG 1288
Db    531 gAspSerLysThrIleGlnLysGlySerGluSerGlyArgGlyArgGlnLys-------- 548

Qy   1289 GTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCTAAGGGAGGTGA 1348
Db    549 ----SerProAlaGlnSerAspSerThrThrGlnArgArgThrValGlyLysLysGlnPr 567

Qy   1349 GATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATGTGAA 1408
Db    567 oLysLysAlaGlu---LysAlaAlaAlaGluGluProArgGlyGlyLeuLys---IleGl 585

Qy   1409 TGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCT-------- 1460
Db    585 uSerGluThrProValAspLeuAlaSerSerMetProSerSerArgHisLysAlaAlaTh 605

Qy   1461 -------------AAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAG 1507
Db    605 rLysGlySerArgLysProAsnIleLysLysGluSerLysSerSerProArg-------- 622

Qy   1508 AGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGA 1567
Db    623 ----ProThrAlaGluLysLysLysTyrLysSerThrSerLysSerSerGlnLysSerAr 641

Qy   1568 AAGGAAGATC-------------------------------------------------- 1577
Db    641 gGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerGluSerLeuProPr 661

Qy   1578 -------------CCTAAATACATTCCACCTCACCTTTCTCCAGATAAG----------- 1613
```

```
                              |||||||||          :::    :::|||   |||
Db        661 oSerSerGlnThrProLysTyrProGluSerAsnArgThrProValLysProSerSerVa 681

Qy       1614 ------------------------AAATGGCTTGGAACTCCTATTGAGGAGATGAGGAG 1648
                                       :::   :::   :::|||:::||||||
Db        681 lGluGluGluAspSerPhePheArgGlnArgMetPheSerProMetGluGluLysGluLe 701

Qy       1649 AATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCACACAGTGAC 1708
               :::                    |||       |||       :::    :::
Db        701 uLeuSer------------------ProLeuSerGluProAspAspArgTyrProLeuIl 715

Qy       1709 TATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCT---AAACCTTTC----------- 1754
                :::::::::|||||       :::       :::|||  ||||||:::
Db        715 eValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLysProTyrLysGluThrGl 735

Qy       1755 ----CCAACACATTTTAAAGATTTGTGGGACAACAAGCATGTT----------------- 1793
                  |||       |||             ||||||
Db        735 uProProLysGlyGluLysLysAsnValProGluLysHisThrArgGluAlaGlnLysGl 755

Qy       1794 ----------AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGG 1843
                        |||:::         :::     :::     ||||||:::|||
Db        755 nAlaSerGluLysValSerAsnLysGlyLysArgLysHisLysAsnGluAspAspAsn-- 774

Qy       1844 TGAGCGAGCTGCAGGCAGCCGG 1865
                  ||||||:::   |||:::
Db        775 ----ArgAlaSerGluSerLys 780
```

RESULT 13
US-10-291-172-197
; Sequence 197, Application US/10291172
; Publication No. US20030228584A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/291,172
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
;   LENGTH: 1163
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-291-172-197

```
Alignment Scores:
Pred. No.:              0.00105       Length:         1163
Score:                  175.50        Matches:        152
Percent Similarity:     33.85%        Conservative:   108
Best Local Similarity:  19.79%        Mismatches:     256
Query Match:            2.42%         Indels:         252
DB:                     15            Gaps:           34

US-09-302-812-1 (1-4070) x US-10-291-172-197 (1-1163)

Qy         59 GGTGGAAGAGGGAAGGCAGGCGTCTGG------ATAGGGCCTGGTTCGGGAGGCTGTCAG 112
              ||||||       ::::::     |||      :::||||||             |||
Db        100 GlyGlySerHisGlnSerSerLysTrpThrProValGlyProAlaProSerThrSerGln 119

Qy        113 AGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGAGGCACCGAGGA 172
              |||:::      |||||||||     |||   ||||||    :::   |||  |||
Db        120 SerGlnLys---ArgSerSerGlyLeuGlnSerGlyHisSerSerGlnArg-ThrSerAl 138

Qy        173 GGGGGCG----------------------------------------------------- 179
               |||:::
Db        138 aGlySerSerSerGlyThrAsnSerSerGlyGlnArgHisAspArgGluSerTyrAsnAs 158

Qy        180 ----------------------------CAGTCCGTCCCTCCCAGGGTTAGTGAATGAGG 211
                                          |||
Db        158 nSerGlySerSerSerArgLysLysGlyGlnHisGlySerGluHisSerLysSerArgSe 178

Qy        212 CTCTACGCCCGGGCTGGCCCGGAGACTCAGTGCTGCGGGTCCCAGCATGAGTGCGGGCCC 271
               |||:::||||||       :::    :::|||:::          |||   |||
Db        178 rSerSerProGlyLysProGlnAlaValSerSerLeuAsnSerSerHisSerArgSerHi 198

Qy        272 CGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCAACTTCTCCGCCGGC 331
               |||   :::      :::|||      :::         :::         |||   ||
Db        198 sGlyAsnAspHisHisSerLysGluHisGlnArgSerLysSerProArgAspProAspAl 218

Qy        332 CGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTCGATTCCAAGGACGC 391
               |      |||:::   |||
Db        218 aAsnTrpAspSerProSer----------------------------------------- 224

Qy        392 TCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTGGGCCGGGCGGACA 451
                           |||||||||   ||||||                         |||||
Db        225 -------------ArgValProPheSerSer------------------------GlyGl 232

Qy        452 GCAC--------------------------CGAGGCAGCGCCACCTCTCTTGTTTTCAA 484
              ||||                                       :::    |||::::::   ||
Db        232 nHisSerThrGlnSerPheProProSerLeuMetSerLysSerAsnSerMetLeuGlnLy 252

Qy        485 ACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAA-------- 536
               |                       ||||||      |||    :::::::::|||
Db        252 sProThrAlaTyrValArgProMetAsp------GlyGlnGluSerMetGluProLysLe 270

Qy        537 ----TCAGAAAGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAG 592
                  ||||||          |||:::    :::    |||:::         |||     ||
Db        270 uSerSerGluHisTyrSerSerGlnSerHisGlyAsnSerMetThrGluLeuLysProSe 290

Qy        593 TTCTGTACAAAAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCA 652
               ||||                            :::   ::::::   ||||||:::   |||||
Db        290 rSer--------------------LysAlaHisLeuThrLysLeuLysIleProSerGl 303

Qy        653 G---------------CTAGGTTTTGATAAGTCACCAGTTGAAAAAGGTACACAGTATTT 697
               |                          |||   |||    |||:::          :::   ||
Db        303 nProLeuAspAlaSerAlaSerGlyAspValSerCysValAsp---------GluIleLe 320

Qy        698 GAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAATGAAGGGCCACACTCAGAACG 757
               ||||:::
Db        320 uLysGlu----------------------------------------------------- 322

Qy        758 GCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTA---------------------- 794
                :::        |||    ||||||    :::|||    :::
Db        323 -MetThrHisSerTrpProProProLeuThrAlaIleHisThrProCysLysThrGluPr 342

Qy        794 ------------------------------------------------------------ 794

Db        342 oSerLysPheProPheProThrLysGluSerGlnGlnSerAsnPheGlyThrGlyGluGl 362

Qy        795 -------------CCAGAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGA 841
                           |||    :::    ||||||    :::         :::|||     |||||
Db        362 nLysArgTyrAsnProSerLysThrSerAsnGlyHisGlnSerLysSerMetLeuLysAs 382

Qy        842 TGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTT---TTGACACA 898
               ||||             ::::::||||||:::|||           |||           |||
Db        382 pAspLeuLysLeuSerSerSerGluAspSerAspGlyGluGlnAspCysAspLysThrMe 402

Qy        899 TGTAAAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAG 958
                  :::                    :::                         |||
Db        402 tProArgSerThrProGlySerAsnSerGluProSerHisHisAsnSerGluGlyAlaAs 422

Qy        959 CCACCAGAAGTGTGGCAAGGCTTGCCATCCT---GCAGAAGCCTGTGCAGGGTGTCAGCA 1015
                :::    :::             :::   |||         :::|||:::      :::|||
Db        422 pAsnSerArgAspAspSerSerHisSerGlySerGluSerSerGlySerAspSe 442
```

```
Qy     1016 GGAGGAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGG 1075
              |||              |||||||||      ||||||:::   :::   :::
Db      442 rGlu---------------SerGluSerSerSerSerAspSerGluAlaAsnGluProSe 457

Qy     1076 TACTGGA------------------------------------------CTGAAAAAT-- 1091
                                                             |||   |||
Db      457 rGlnSerAlaSerProGluProGluProProProThrAsnLysTrpGlnLeuAspAsnTr 477

Qy     1092 -GCCAACAGATTGAATAGACAA-----------------------GAAAGTAGTCTAGG 1126
                 |||:::::|||                             :::|||:::::
Db      477 pLeuAsnLysValAsnProHisLysValSerProAlaSerSerValAspSerAsnIlePr 497

Qy     1127 AAATTCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTC 1186
              :::|||      ::::::|||||        :::                    ::
Db      497 oSerSerGlnGlyTyrLysLysGluGlyArgGluGln------------------GlyTh 511

Qy     1187 CAAAAATAGTTGTCAGGATTCAGAAGCAGATGAAGAGACAAGT---------CCAGGTTT 1237
             :    ||||||      |||:::           :::|||||||||      ||||||
Db      511 rGlyAsnSerTyrThrAspThrSerGlyProLysGluThrSerSerAlaThrProGlyAr 531

Qy     1238 TGATGAA---------CAGGAAGATAGCAGTTCTGCTCAAACAGCAAATAAACCTTCAAG 1288
              |||             |||:::   |||   |||   :::         |||
Db      531 gAspSerLysThrIleGlnLysGlySerGluSerGlyArgGlyArgGlnLys-------- 548

Qy     1289 GTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCTAAGGGAGGTGA 1348
                 |||    ::::::|||:::        ::::::|||:::         |||
Db      549 ----SerProAlaGlnSerAspSerThrThrGlnArgArgThrValGlyLysLysGlnPr 567

Qy     1349 GATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATGAATGATGTGAA 1408
               :::           :::          |||    |||    |||:::     :::
Db      567 oLysLysAlaGlu---LysAlaAlaAlaGluGluProArgGlyGlyLeuLys---IleGl 585

Qy     1409 TGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTCT-------- 1460
             ::::::   |||            :::|||                |||:::
Db      585 uSerGluThrProValAspLeuAlaSerSerMetProSerSerArgHisLysAlaAlaTh 605

Qy     1461 ------------AAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAG 1507
                         |||   :::   :::|||:::||||||    :::
Db      605 rLysGlySerArgLysProAsnIleLysLysGluSerLysSerSerProArg-------- 622

Qy     1508 AGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGA 1567
                 |||   ||||||   |||:::    :::                  |||::::::
Db      623 ----ProThrAlaGluLysLysLysTyrLysSerThrSerLysSerSerGlnLysSerAr 641

Qy     1568 AAGGAAGATC-------------------------------------------------- 1577
                   |||
Db      641 gGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerGluSerLeuProPr 661

Qy     1578 ------------CCTAAATACATTCCACCTCACCTTTCTCCAGATAAG----------- 1613
                        |||||||||        :::   :::|||   |||
Db      661 oSerSerGlnThrProLysTyrProGluSerAsnArgThrProValLysProSerSerVa 681

Qy     1614 ------------------------AAATGGCTTGGAACTCCTATTGAGGAGATGAGGAG 1648
                                    :::   :::   :::|||:::||||||
Db      681 lGluGluGluAspSerPhePheArgGlnArgMetPheSerProMetGluGluLysGluLe 701

Qy     1649 AATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCACACAGTGAC 1708
              :::                     |||        |||         :::   :::
Db      701 uLeuSer-----------------ProLeuSerGluProAspAspArgTyrProLeuIl 715

Qy     1709 TATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCT---AAACCTTTC----------- 1754
              :::::::::||||||        :::        :::|||   ||||||:::
Db      715 eValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLysProTyrLysGluThrGl 735

Qy     1755 ----CCAACACATTTTAAAGATTTGTGGGACAACAAGCATGTT----------------- 1793
                |||        |||                   ||||||
Db      735 uProProLysGlyGluLysLysAsnValProGluLysHisThrArgGluAlaGlnLysGl 755

Qy     1794 ----------AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGG 1843
                      |||:::         :::       :::      ||||||:::|||
Db      755 nAlaSerGluLysValSerAsnLysGlyLysArgLysHisLysAsnGluAspAspAsn-- 774

Qy     1844 TGAGCGAGCTGCAGGCAGCCGG 1865
```

```
                   ||||||:::   |||:::
Db        775 ----ArgAlaSerGluSerLys 780
```

RESULT 14
US-10-287-218-17
; Sequence 17, Application US/10287218
; Publication No. US20030198975A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: AZIMZAI, Yalda; AU-YOUNG, Janice K.
; APPLICANT: BATRA, Sajeev; BAUGHN, Mariah R.
; APPLICANT: BECHA, Shanya D.; BOROWSKY, Mark L.
; APPLICANT: BUFORD, Neil; DING, Li
; APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.
; APPLICANT: GANDHI, Ameena R.; GIETZEN, Kimberly J.
; APPLICANT: GRIFFIN, Jennifer A.; HAFALIA, April J.A.
; APPLICANT: HONCHELL, Cynthia D.; LAL, Preeti G.
; APPLICANT: LEE, Soo Yeun; LU, Dyung Aina M.
; APPLICANT: ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
; APPLICANT: REDDY, Roopa; SANJANWALA, Madhu, M.
; APPLICANT: TANG, Y. Tom; WALIA, Narinder K.
; APPLICANT: WANG, Yu-mei, E.; WARREN, Bridget A.
; APPLICANT: XU, Yuming; YANG, Junming
; APPLICANT: YAO, Monique G.; YUE, Henry
; APPLICANT: ZEBARJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PI-0417 USA
; CURRENT APPLICATION NUMBER: US/10/287,218
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: PCT/US02/11152
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US 60/349,705
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/295,263
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/295,340
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/293,727
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/291,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/291,662
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: US 60/287,228
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/286,820
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/283,294
; PRIOR FILING DATE: 2001-04-11
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 17
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030198975A1 5734806CD1
US-10-287-218-17

Alignment Scores:			
Pred. No.:	0.00135	Length:	1462
Score:	174.50	Matches:	158
Percent Similarity:	33.84%	Conservative:	110
Best Local Similarity:	19.95%	Mismatches:	298
Query Match:	2.41%	Indels:	226
DB:	14	Gaps:	32

US-09-302-812-1 (1-4070) x US-10-287-218-17 (1-1462)

```
Qy      113 AGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGGCATGGTGCCGGGAGGCACCGAGGA 172
              :::   |||   |||   |||   |||    :::  |||                  ||||||
```

```
Db     511 ThrHisSerSerArgAspSerTyrArgSerLysSerHisSerGlnSerTyrSerArgGly 530
Qy     173 GGGGGCGCAGTCCGTCCCTCCCAGGGTTAGTGAATGAGGCTCTACGCCCGGGCTGGCCCG 232
              ::::::|||          |||   :::        |||:::        :::||
Db     531 -SerSerArgSer---------ArgThrAlaSerLysSerSerSerHisSerArgSerAr 547
Qy     233 GAGACTCAGTGCTGCGGGTCCCAGCATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAA 292
           |        |||          |||   |||            :::      :::||
Db     547 gSerLysSerArg---------SerSerSerLysSerGlyHisArgLysArgAlaSerLy 564
Qy     293 GCGACCCCGCTGGGACGCCGCT-----------------------GCAACTTCTCCGCC 328
           |    ||||||     |||:::                          |||   |||
Db     564 sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProLe 584
Qy     329 GGCCGCCTCGGACGCCCGGAGC--------------------TTCCCCGGCAGGCAGAG 367
              |||:::   |||::::::
Db     584 uArgAlaThrMetAlaGlnAsnGluAsnValValValGlnProValValAlaGluAsnIl 604
Qy     368 GCGCGTCCTCGATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGG 427
              |||:::            |||:::|||            |||      |||
Db     604 eProValIleProLeuSerAspSerProProProSerArgTrpLysProGlyGlnLysPr 624
Qy     428 CTGC------GCCCTGGGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTT 481
                       :::      |||   :::   :::            |||   |||:::
Db     624 oTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHisLeuLeuPr 644
Qy     482 CAAACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGA 541
              |||   |||      |||      :::   |||      :::|||      |||
Db     644 oIleGlnSerThrTyr---SerLeuAlaAsnIleLys------GluThrGlySerSerSe 661
Qy     542 AAGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTA-- 599
            |||   |||   :::|||   |||:::         ::::::|||         |||
Db     661 rSerTyrHisLysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSe 681
Qy     600 ------------------------------------------CAAAAAGATAACTTTTA 616
                                                        :::          ||
Db     681 rAspArgSerSerGluSerSerProArgSerArgSerArgSerSerArgSerArgSerTy 701
Qy     617 TCAACATAACATGGAAAAATTAGAAAATGTT---TCTCAGCTAGGTTTTGATAAGTCACC 673
           |        :::      :::      ::::::     |||            :::|||||
Db     701 rSerArgSerTyrThrArgSerArgSerLeuAlaSerSerHisSerArgSerArgSerPr 721
Qy     674 AGTTGAAAAAGGT---------ACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTG 724
           |        :::                  :::|||            |||   :::
Db     721 oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 741
Qy     725 TAAGTGG------------CAGAATGAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGA 772
              :::                     ::::::|||   ::::::|||   |||   |||
Db     741 rSerTyrThrSerIleSerSerAspAspGlyArgArgAlaLysArgArgLeuArgSerSe 761
Qy     773 ACCTCCAGCGGTAACTCTGGTACCAGAGCAGTTCAGTAATGCTAATGTC----------- 821
                          ::::::      ::::::   |||:::::::::
Db     761 rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerGluLysThrLeuHi 781
Qy     822 ---------------------GATCAGTCGTCCCCAAAGGATGATCACAGTGACACAAA 859
                                 |||:::||||||            :::|||::::::
Db     781 sSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSerGluSerAr 801
Qy     860 TAGT------GAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAA 913
             |||      :::   |||   |||::::::|||         :::   |||:::
Db     801 gSerSerLeuAspTyrSerSerAspSerGluGln------SerSerValGlnAlaThrGl 819
Qy     914 TGCAAAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGG 973
             :::   |||         |||   :::||||||:::         |||:::|||   |||
Db     819 nSerAlaGlnGluLysGlu---LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 838
Qy     974 CAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGT 1033
             |||                                    :::|||::::::
Db     838 uLys--------------------------AsnArgGlyGluGluLysSerLysSerGl 849
Qy    1034 GTCCGAGAGCCCCTTGTCGGAC------ACTGGCTCTGAGGATGTTGGTACTGGACTGAA 1087
              |||   |||   |||         |||      |||::::::          |||::
Db     849 uArgGluCysProHisSerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuAr 869

Qy    1088 AAATGCCAACAGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAA 1147
            :|||    ::::::       |||:::               |||         :::
Db     869 gAsnGlySerLysProLysArgLys---------------AsnTyrAlaGlySerLysTr 884
Qy    1148 AGAAAGTGAACCTGAGTCACCAATGGATGTA---GATAATTCCAAAAATAGTTGTCAGGA 1204
            :::||||||      |||         ||||||      |||||||||||
Db     884 pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 904
Qy    1205 TTCAGAAGCAGATGAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGC 1264
             |||      ||||||                 |||   :::   :::   |||
Db     904 oSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluSerGluValSerGluIl 924
Qy    1265 TCAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAA 1324
                          ||||||:::
Db     924 eHisIleLysValLysProThr-----------------------------------Th 932
Qy    1325 GCGGTCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAG 1384
             :::||||||                                               ::
Db     932 rLysSerSer---------------------------------------------Th 936
Qy    1385 TCGAGCTGGAATGAATGATGTGAAT---------------GCCAAACGACCTGGAAGTAC 1429
           :       :::   |||   |||              :::|||:::         |||||
Db     936 rAsnThrSerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerTh 956
Qy    1430 TTCTAGCCTGAATGTAGAGTGCAGAAATTCTAAG------------------------- 1463
            ||||                  |||   ||||||:::
Db     956 rSerAspSerGluGlySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHi 976
Qy    1464 -CAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAAGCAGA 1522
             :::||||||   |||::::::            :::|||         |||   ::
Db     976 sLysHisGlySerLysGluAsnLeuLysArgGluHis---------ThrLysLysValLy 993
Qy    1523 GGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATCCCTAA 1582
             :::|||   |||   :::   :::   ||||||:::         :::||||||
Db     993 sGluLysLeuLysGlyLysLysAspLysLysHisLysAlaProLysArgLysGlnAlaPh 1013
Qy    1583 ATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGAT 1642
             :::   |||
Db    1013 eHisTrpGlnPro--------------------------------------------- 1017
Qy    1643 GAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCACAC 1702
Db    1017 ------------------------------------------------------------ 1017
Qy    1703 AGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCAACACA 1762
                                                   |||      |||
Db    1018 -------------------------------------ProLeuGluPheGlyGluGl 1024
Qy    1763 TTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAACTTGTA 1822
                         |||:::|||   |||                  |||:::   :::
Db    1024 uGluGluGluIleAspAspLysGlnValThrGlnGluSerLysGluLysLysVal-- 1043
Qy    1823 CCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATTCAGAC 1882
                        ||||||   |||
Db    1044 ----------SerGluAsnAsnGlu--------------------------------- 1048
Qy    1883 TGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAA 1942
                                           :::||||||   |||||||||
Db    1049 ------------------------------ThrIleLysAspAsnIleLeuLysThrGl 1058
Qy    1943 TGTGGCA-----------TATTCTAAGAAATGGACTTTACAGCTTTGATTGATTTCTG 1990
              :::             |||   |||   |||               :::
Db    1058 uLysSerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLe 1078
Qy    1991 GGATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATAT 2050
             |||:::      ::::::            :::   :::   |||
Db    1078 uAspGlnPheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrGl 1098
Qy    2051 GGTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACA 2110
              :::|||   ||||||   ||||||                                ||
Db    1098 uGluAsnValAla---CysLeuGlnAsnIle--------------------------Gl 1108
```

```
Qy        2111 GAAGATGAATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGC 2170
Db        1108 nHisValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMe 1128

Qy        2171 TTTCTTCTGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGT----- 2225
Db        1128 tGluIleCysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLe 1147

Qy        2226 ------------------TATCCAGATATTAAC 2240
Db        1147 uGlyAsnAlaArgLeuAspThrProAspIleAsn 1158
```

RESULT 15
US-09-932-257A-3
; Sequence 3, Application US/09932257A
; Publication No. US20030039658A1
; GENERAL INFORMATION:
; APPLICANT: Estable, Mario
; APPLICANT: Roeder, Robert
; TITLE OF INVENTION: MCEF, A No. US20030039658A1el Transcription Factor
; FILE REFERENCE: 600-1-269N
; CURRENT APPLICATION NUMBER: US/09/932,257A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/226,340
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/226,339
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1086)..(1086)
; OTHER INFORMATION: x is any amino acid
US-09-932-257A-3

Alignment Scores:

Pred. No.:	0.00147	Length:	1163
Score:	173.50	Matches:	150
Percent Similarity:	31.91%	Conservative:	96
Best Local Similarity:	19.46%	Mismatches:	267
Query Match:	2.40%	Indels:	258
DB:	10	Gaps:	33

US-09-302-812-1 (1-4070) x US-09-932-257A-3 (1-1163)

```
Qy          59 GGTGGAAGAGGGAAGGCAGGCGTCTGG------ATAGGGCCTGGTTCGGGAGGCTGTCAG 112
Db         100 GlyGlySerHisGlnSerSerLysTrpThrProValGlyProAlaProSerThrSerGln 119

Qy         113 AGCAGGAGCTGCAGAAGCAGTCAGCGGCAGAGGGGCATGGTGCCGGGAGGCACCGAGGA 172
Db         120 SerGlnLys---ArgSerSerGlyLeuGlnSerGlyHisSerSerGlnArg-ThrSerAl 138

Qy         173 GGGGGCG----------------------------------------------------- 179
Db         138 aGlySerSerSerGlyThrAsnSerSerGlyGlnArgHisAspArgGluSerTyrAsnAs 158

Qy         180 ---------------------------CAGTCCGTCCCTCCCAGGGTTAGTGAATGAGG 211
Db         158 nSerGlySerSerSerArgLysLysGlyGlnHisGlySerGluHisSerLysSerArgSe 178

Qy         212 CTCTACGCCCGGGCTGGCCCGGAGACTCAGTGCTGCGGGTCCCAGCATGAGTGCGGGCCC 271
Db         178 rSerSerProGlyLysProGlnAlaValSerSerLeuAsnSerSerHisSerArgSerHi 198

Qy         272 CGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCAACTTCTCCGCCGGC 331
Db         198 sGlyAsnAspHisHisSerLysGluHisGlnArgSerLysSerProArgAspProAspAl 218

Qy         332 CGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTCGATTCCAAGGACGC 391
Db         218 aAsnTrpAspSerProSer----------------------------------------- 224

Qy         392 TCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTGGGCCGGGCGGGACA 451
Db         225 -------------ArgValProPheSerSer-----------------------GlyGl 232

Qy         452 GCAC---------------------------CGAGGCAGCGCCACCTCTCTTGTTTTCAA 484
Db         232 nHisSerThrGlnSerPheProProSerLeuMetSerLysSerAsnSerMetLeuGlnLy 252

Qy         485 ACAGAAGACTATAACCAGTTGGATGGACACTAAAGGAATC------AAGACAGTTGAATC 538
Db         252 sProThrAlaTyrValArgProMetAspGlyGlnGlySerMetGluProLysLeuSerSe 272

Qy         539 AGAAAGTTTGCATAGTAAAGAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGT 598
Db         272 rGluHisTyrSerSerGlnSerHisGlyAsnSerMetThrGluLeuLysProSerSer-- 291

Qy         599 ACAAAAAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATGTTTCTCAG----- 653
Db         292 ------------------LysAlaHisLeuThrLysLeuLysIleProSerGlnProLe 305

Qy         654 ----------CTAGGTTTTGATAAGTCACCAGTTGAAAAAGGTACACAGTATTTGAAGCA 703
Db         305 uAspAlaSerAlaSerGlyAspValSerCysValAsp---------GluIleLeuLysGl 322

Qy         704 GCATCAGACTGCGGCTATGTGTAAGTGGCAGAATGAAGGGCCACACTCAGAACGGCTTTT 763
Db         322 u-----------------------------------------------------MetTh 324

Qy         764 GGAAAGTGAACCTCCAGCGGTAACTCTGGTA---------------------CCAGAGCA 802
Db         324 rHisSerTrpProProProLeuThrAlaIleHisThrProCysLysThrGluProSerLy 344

Qy         803 GTTC--------------------------AGTAATGCTAATGTCGATCAGTCGTCCCC 835
Db         344 sPheProPheProThrLysGluSerGlnGlnSerAsnPheGlyIleGlyGluGlnLysAr 364

Qy         836 AAAGGATGATCACAGTGACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGAC 895
Db         364 gTyrAsnProSerLysThrSerAsnGlyHisGlnSerLys------SerMetLeuLysAs 382

Qy         896 ACATGTAAAGCTTGCGAATGCAAAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAG 955
Db         382 pGlyLeuLysLeuSerSerSer------------GluAspSerAspGly----------- 394

Qy         956 AAGCCACCAGAAGTGTGGCAAGGCTTGC-------------------------------- 983
Db         395 ----GluGlnAspCysAspLysThrMetProArgSerThrProGlySerAsnSerGluPr 413

Qy         984 -------CATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTC 1036
Db         413 oSerHisHisAsnSerGlu---------GlyAlaAspAsnSerArgAspAspSerSerSe 430

Qy        1037 CGAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAA 1096
Db         430 rHisSerGlySerGluSerSerSerGlySerAspSerGluSerGlySerSerSerSerAs 450

Qy        1097 CAGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGA 1156
Db         450 p--------------------SerGluAlaAsnGluProSerGlnSerAlaSerProGl 463

Qy        1157 ACCTGAGTCACCA------------ATGGATGTAGATAAT-------------------- 1184
Db         463 uProGluProProProThrAsnLysTrpGlnLeuAspAsnTrpLeuAsnLysValAsnPr 483

Qy        1184 ------------------------------------------------------------ 1184
Db         483 oHisLysValSerProAlaSerSerValAspSerAsnIleProSerSerGlnGlyTyrLy 503

Qy        1185 ----------------------TCCAAAAATAGTTGTCAGGATTCAGAAGCAGATGAAGA 1222
```

```
                                 :::   ||||||       |||:::          :::||
Db        503 sLysGluGlyArgGluGlnGlyThrGlyAsnSerTyrThrAspThrSerGlyProLysGl 523

Qy       1223 GACAAGT---------CCAGGTTTTGATGAA---------CAGGAAGATAGCAGTTCTGC 1264
              |||||||         ||||||  |||               |||:::  |||   |||
Db        523 uThrSerSerAlaThrProGlyArgAspSerLysThrIleGlnLysGlySerGluSerGl 543

Qy       1265 TCAAACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAA 1324
                 :::          |||                  |||   ::::::|||:::     :::::
Db        543 yArgGlyArgGlnLys------------SerProAlaGlnSerAspSerThrThrGlnAr 559

Qy       1325 GCGGTCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAG 1384
              :|||:::        |||                :::          :::              |||
Db        559 gArgThrValGlyLysLysGlnProLysLysAlaGlu---LysAlaAlaAlaGluGluPr 578

Qy       1385 TCGAGCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGT 1444
                 |||   |||:::           :::                     ||||||:::
Db        578 oArgGlyGlyLeuLysIleGluSerGluThrProValAspLeuAlaSerSerMetProSe 598

Qy       1445 AGAGTGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCAT 1504
                             :::           |||   :::          |||           |||
Db        598 rSerArgHisLysAlaAlaThrLysGlySerArgLysProAsnIleLysLysGluPheLy 618

Qy       1505 GAGAGTGCCCAAA---GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAG 1561
                       |||:::         :::|||:::|||    :::          |||         ::
Db        618 sSerSerProArgProThrAlaGluLysLysLysTyrLysSerThrSerLysSerSerGl 638

Qy       1562 AACAGAAAGGAAGATC-------------------------------------------- 1577
              :        |||:::|||
Db        638 nLysSerArgGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerGluSe 658

Qy       1578 ---------------------CCTAAATACATTCCACCTCACCTTTCTCCAGATAAG-- 1613
                                   |||||||||           :::     :::|||     |||
Db        658 rLeuProProSerSerGlnThrProLysTyrProGluSerAsnArgThrProValLysPr 678

Qy       1614 ---------------------------------AAATGGCTTGGAACTCCTATTGAGGA 1639
                                               :::     :::     :::|||:::|||||
Db        678 oSerSerValGluGluGluAspSerPhePheArgGlnArgMetPheSerProMetGluGl 698

Qy       1640 GATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCA 1699
              |         :::                        |||          |||          ::
Db        698 uLysGluLeuLeuSer------------------ProLeuSerGluProAspAspArgTy 712

Qy       1700 CACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCT---AAACCTTTC-- 1754
              :   :::   :::::::::||||||       :::      :::|||   ||||||:::
Db        712 rProLeuIleValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLysProTyrLy 732

Qy       1755 -------------CCAACACATTTTAAAGATTTGTGGGACAACAAGCATGTT-------- 1793
                           |||          |||                ||||||
Db        732 sGluThrGluProProLysGlyGluLysLysAsnValProGluLysHisThrArgGluAl 752

Qy       1794 ------------------AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGA 1834
                                |||:::          :::       :::         |||||
Db        752 aGlnLysGlnAlaSerGluLysValSerAsnLysGlyLysArgLysHisLysAsnGluAs 772

Qy       1835 TGAGAATGGTGAGCGAGCTGCAGGCAGCCGG 1865
              |:::|||         ||||||:::    |||:::
Db        772 pAspAsn------ArgAlaSerGluSerLys 780
```

Search completed: May 26, 2004, 18:34:40
Job time : 252.229 secs

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 26, 2004, 16:27:48 ; Search time 34.05 Seconds
(without alignments)
12341.705 Million cell updates/sec

Title: US-09-302-812-1
Perfect score: 7242
Sequence: 1 accggaaagtgaacgaagcc..........aaattttcattaacaaaaaa 4070

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 77882

9511477
9511507

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150054_5717/
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TF
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -M
-USER=6333148_@CGN_1_1_131_@runat_26052004_150054_5717 -NCPU
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5184	71.6	977	4	US-09-302-812-2	Sequence 2, Appli
2	5184	71.6	977	4	US-09-511-477-2	Sequence 2, Appli
3	5184	71.6	977	4	US-09-511-507-2	Sequence 2, Appli
4	4635.5	64.0	976	4	US-09-302-812-4	Sequence 4, Appli
5	4635.5	64.0	976	4	US-09-511-477-4	Sequence 4, Appli
6	4635.5	64.0	976	4	US-09-511-507-4	Sequence 4, Appli
7	4332.5	59.8	968	4	US-09-302-812-6	Sequence 6, Appli
8	4332.5	59.8	968	4	US-09-511-477-6	Sequence 6, Appli
9	4332.5	59.8	968	4	US-09-511-507-6	Sequence 6, Appli
10	1063.5	14.7	768	4	US-09-302-812-8	Sequence 8, Appli
11	1063.5	14.7	768	4	US-09-511-477-8	Sequence 8, Appli
12	1063.5	14.7	768	4	US-09-511-507-8	Sequence 8, Appli
13	465	6.4	726	4	US-09-302-812-10	Sequence 10, Appl
14	465	6.4	726	4	US-09-511-477-10	Sequence 10, Appl
15	465	6.4	726	4	US-09-511-507-10	Sequence 10, Appl
16	173	2.4	1187	1	US-08-320-559-28	Sequence 28, Appl
17	173	2.4	1187	3	US-08-545-860D-28	Sequence 28, Appl
18	173	2.4	1187	5	PCT-US94-04496-28	Sequence 28, Appl
19	173	2.4	1210	1	US-08-320-559-26	Sequence 26, Appl
20	173	2.4	1210	3	US-08-545-860D-26	Sequence 26, Appl
21	173	2.4	1210	5	PCT-US94-04496-26	Sequence 26, Appl
22	169.5	2.3	913	4	US-08-971-089-4	Sequence 4, Appli
23	168	2.3	778	6	5198347-4	Patent No. 5198347
24	167	2.3	31	4	US-09-302-812-11	Sequence 11, Appl
25	167	2.3	31	4	US-09-511-477-11	Sequence 11, Appl
26	167	2.3	31	4	US-09-511-507-11	Sequence 11, Appl
27	164.5	2.3	1804	4	US-09-362-336A-4	Sequence 4, Appli
28	162	2.2	1404	4	US-08-801-308-1	Sequence 1, Appli
29	161	2.2	951	4	US-09-125-635-8	Sequence 8, Appli
30	161	2.2	1420	4	US-09-125-635-4	Sequence 4, Appli
31	154.5	2.1	1008	4	US-09-252-991A-29419	Sequence 29419, A
32	154	2.1	703	3	US-08-910-925-4	Sequence 4, Appli
33	153.5	2.1	8991	4	US-08-714-741-32	Sequence 32, Appl
		2.1	914	1	US-08-484-105-2	Sequence 2, Appli
		2.1	914	1	US-08-484-106-2	Sequence 2, Appli
		2.1	1783	4	US-09-362-336A-2	Sequence 2, Appli
		2.1	1481	2	US-08-616-844-40	Sequence 40, Appl
		2.1	1481	2	US-08-599-654-40	Sequence 40, Appl
		2.1	1481	3	US-08-944-868A-40	Sequence 40, Appl
		2.1	1481	3	US-08-944-423A-40	Sequence 40, Appl
		2.1	1481	3	US-08-944-496-40	Sequence 40, Appl
		2.1	29	4	US-09-302-812-12	Sequence 12, Appl
		2.1	29	4	US-09-511-477-12	Sequence 12, Appl
		2.1	29	4	US-09-511-507-12	Sequence 12, Appl
		2.0	26	4	US-09-302-812-19	Sequence 19, Appl

ALIGNMENTS

```
..plication US/09302812B
,33148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-302-812-2

Alignment Scores:
Pred. No.:                0            Length:          977
Score:                    5184.00      Matches:         977
Percent Similarity:       100.00%      Conservative:    0
Best Local Similarity:    100.00%      Mismatches:      0
Query Match:              71.58%       Indels:          0
DB:                       4            Gaps:            0

US-09-302-812-1 (1-4070) x US-09-302-812-2 (1-977)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy        918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy        978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy       1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy       1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy       1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy       1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy       1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy       1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy       1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy       1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy       1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy       1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy       1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy       1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy       1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy       1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy       1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy       1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy       1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy       2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy       2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740
```

```
Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGGGCC 3188
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977
```

RESULT 2
US-09-511-477-2
; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-2

```
Alignment Scores:
Pred. No.:                 0              Length:          977
Score:                     5184.00        Matches:         977
Percent Similarity:        100.00%        Conservative:    0
Best Local Similarity:     100.00%        Mismatches:      0
Query Match:               71.58%         Indels:          0
DB:                        4              Gaps:            0

US-09-302-812-1 (1-4070) x US-09-511-477-2 (1-977)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy        918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy        978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy       1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy       1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy       1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy       1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy       1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy       1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy       1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy       1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy       1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy       1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy       1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy       1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy       1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy       1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy       1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy       1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy       1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy       2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy       2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGGGCC 3188
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977
```

RESULT 3
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH

```
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-511-507-2

Alignment Scores:
Pred. No.:              0             Length:        977
Score:                  5184.00       Matches:       977
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            71.58%        Indels:        0
DB:                     4             Gaps:          0

US-09-302-812-1 (1-4070) x US-09-511-507-2 (1-977)

Qy      258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy      318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy      378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy      438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy      498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy      558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy      618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy      678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy      738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy      798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy      858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy      918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy      978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy     1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy     1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy     1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy     1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy     1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy     1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy     1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy     1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy     1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy     1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy     1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy     1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy     1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy     1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy     1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy     1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy     1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy     2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620
```

```
Qy       2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy       2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGGGCC 3188
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977
```

```
RESULT 4
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-302-812-4

Alignment Scores:
Pred. No.:               0              Length:         976
Score:                   4635.50        Matches:        874
Percent Similarity:      92.83%         Conservative:   32
Best Local Similarity:   89.55%         Mismatches:     69
Query Match:             64.01%         Indels:         1
DB:                      4              Gaps:           1

US-09-302-812-1 (1-4070) x US-09-302-812-4 (1-976)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              |||:::|||||||||||||||||||||||||||||          ||||||   ||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||   ||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              |||  |||||||||  |||||||||||||||||||||||||||         |||
Db         40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTCAAACAGAAGACTATA 497
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              ||||||||||||||||||||||||||||||||||   |||||||||||||||   ||||||
Db         80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              |||||||||:::|||||||||   |||||||||||||||||    |||||||||
Db        120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||   |||:::|||:::||||||||||||||||||         ||||||||||||||||
Db        160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              ||||||||||||||||||||||:::|||:::|||         ||||||||||||||||||
```

```
Db    180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199
Qy    858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
Db    200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
Qy    918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db    220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239
Qy    978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db    240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259
Qy   1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db    260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279
Qy   1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db    280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299
Qy   1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db    300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319
Qy   1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db    320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339
Qy   1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db    340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359
Qy   1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db    360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379
Qy   1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db    380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399
Qy   1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db    400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419
Qy   1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db    420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439
Qy   1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db    440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459
Qy   1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db    460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479
Qy   1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db    480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499
Qy   1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db    500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519
Qy   1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db    520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539
Qy   1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db    540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559
```

```
Qy   1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db    560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579
Qy   1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db    580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599
Qy   2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db    600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619
Qy   2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db    620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639
Qy   2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db    640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659
Qy   2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db    660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679
Qy   2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db    680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699
Qy   2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db    700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719
Qy   2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db    720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739
Qy   2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db    740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759
Qy   2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db    760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779
Qy   2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db    780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799
Qy   2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
Db    800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819
Qy   2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
Db    820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839
Qy   2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
Db    840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859
Qy   2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
Db    860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879
Qy   2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
Db    880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899
Qy   2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
Db    900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919
```

```
Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
              ||||||||||||   :::|||   :::       ||||||||||||||
Db        960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975
```

RESULT 5
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-477-4

```
Alignment Scores:
Pred. No.:              0           Length:         976
Score:                  4635.50     Matches:        874
Percent Similarity:     92.83%      Conservative:   32
Best Local Similarity:  89.55%      Mismatches:     69
Query Match:            64.01%      Indels:         1
DB:                     4           Gaps:           1

US-09-302-812-1 (1-4070) x US-09-511-477-4 (1-976)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              |||:::||||||||||||||||||||||||||||||||||        ||||||   ||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||   |||||||||||||||||||||||||||||||    ||||||||||||||||||
Db         21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              |||    |||||||||   ||||||||||||||||||||||||||||||         |||
Db         40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              |||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              ||||||||:::|||||||||     ||||||||||||||||       |||||||||
Db        120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||   ||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||   |||:::|||:::|||||||||||||||||||         ||||||||||||||
Db        160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              |||||||||||||||||||||:::|||:::|||        ||||||||||||||||||
Db        180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
              :::|||||||||:::||||||||||||||||||||||||||    |||||||||||||||||
Db        200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219

Qy        918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
              |||||||||   ||||||||||       |||||||||:::|||||||||||||||   |||
Db        220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy        978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
              :::||||||||||   |||   ||||||   ||||||||||:::|||   ||||||||||
Db        240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy       1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
              :::||||||||||||||||   ||||||||||||||||||||||||||||    ||||||   |||
Db        260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279

Qy       1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
              :::|||   ||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy       1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy       1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
              ||||||||||||||||||||||||||||||||||||   ||||||:::||||||||||||
Db        320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy       1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
              ||||||||||||||||||   |||:::||||||   |||   |||||||||   |||
Db        340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy       1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
              |||||||||||||:::||||||||||||||||||||||||||||||||||   ||||||
Db        360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy       1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
              ||||||:::|||||||||   ||||||:::   ||||||||||||||||||||||||||||
Db        380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy       1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
              ||||||||||||||:::||||||||||||||||||||||||   ||||||:::||||||
Db        400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419

Qy       1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
              |||||||||:::||||||||||||   |||   |||||||||||||||||||||||||||
Db        420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy       1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
              ||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459

Qy       1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy       1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499
```

```
Qy       1758 ACACATTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
              |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy       1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
              |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy       1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
              ||||||||||||||||||:::      |||||||||||||||||||||||||||||||||
Db        540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy       1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy       1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy       2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy       2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy       2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
              |||||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db        700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799

Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              ||||||:::||||||||||||||||   |||||||||||||||   :::||||||   |||
Db        800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db        880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
              |||||||||||    :::|||    :::        ||||||||||||
Db        960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975
```

RESULT 6
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-507-4

```
Alignment Scores:
Pred. No.:                0              Length:         976
Score:                    4635.50        Matches:        874
Percent Similarity:       92.83%         Conservative:   32
Best Local Similarity:    89.55%         Mismatches:     69
Query Match:              64.01%         Indels:         1
DB:                       4              Gaps:           1

US-09-302-812-1 (1-4070) x US-09-511-507-4 (1-976)

Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              |||:::|||||||||||||||||||||||||||||         ||||||   ||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||   ||||||||||||||||||||||||||   ||||||||||||||||||
Db         21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              |||   |||||||||   ||||||||||||||||||||||||||          |||
Db         40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79

Qy     498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
Db      80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99

Qy     558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
Db     100 GluAsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119

Qy     618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
Db     120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139

Qy     678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
Db     140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159

Qy     738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db     160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179

Qy     798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
Db     180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199

Qy     858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
Db     200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219

Qy     918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db     220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239

Qy     978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db     240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259

Qy    1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db     260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279

Qy    1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db     280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

Qy    1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db     300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319

Qy    1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db     320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339

Qy    1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db     340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359

Qy    1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db     360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379

Qy    1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db     380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399

Qy    1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db     400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419

Qy    1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db     420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

Qy    1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db     440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459

Qy    1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db     460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479

Qy    1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db     480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499

Qy    1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db     500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519

Qy    1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db     520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539

Qy    1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db     540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

Qy    1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db     560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579

Qy    1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db     580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599

Qy    2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db     600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619

Qy    2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db     620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639

Qy    2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db     640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659

Qy    2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db     660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679

Qy    2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db     680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699

Qy    2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db     700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719

Qy    2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db     720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739

Qy    2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db     740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759

Qy    2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db     760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779

Qy    2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db     780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799
```

```
Qy       2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
              |||||:::||||||||||||||  ||||||||||||||    :::||||||   |||
Db        800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819

Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839

Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879

Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 899

Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939

Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959

Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
              |||||||||||   :::|||   :::      ||||||||||||||
Db        960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975
```

RESULT 7
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-302-812-6

```
Alignment Scores:
Pred. No.:              0           Length:         968
Score:                  4332.50     Matches:        820
Percent Similarity:     89.86%      Conservative:   57
Best Local Similarity:  84.02%      Mismatches:     90
Query Match:            59.82%      Indels:         9
DB:                     4           Gaps:           5
```

US-09-302-812-1 (1-4070) x US-09-302-812-6 (1-968)

```
Qy        258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
              |||||||||||||||||| |||||||||||||||       ||||||  ||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19

Qy        318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
              ||||||   |||   |||||||||:::||||||||||||||||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
              |||   ||||||||||||||||||||||||||||||||||||||||       |||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
              ||||||||||||   |||||||||:::|||||||||   |||||||||||||||||||
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
              |||:::||||||||||||||||   ||||||   |||||||||         ||||||
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
              ||||||||||||||||||   :::||||||||||||||||||||||||||||||||||||
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
              |||   :::||||||||||||||||||   ||||||      ||||||||||||
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
              ||||||   :::|||||||||   |||   ||||||||||::::::||||||||||||||
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
              ||||||   |||:::|||:::||||||   ||||||||||||||||   |||   :::|||
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
              :::|||   ||||||||||||:::   ||||||         ||||||||||||||||||
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
              :::   ||||||   ||||||||||||||||||||||||   :::||||||||||||
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
              |||   |||:::   |||   |||            |||||||||:::   ||||||
Db        217 LysProThrValGlyAspGlyGln---------AlaArgSerAsnCysLysCysSerGly 233

Qy        978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
              :::            :::   |||   ||||||||||||||||||   ||||||:::
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy       1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
              ||||||||||||||||||   |||:::||||||:::|||||||||   ||||||   |||
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
              :::|||      ||||||||||||||||||||||:::|||||||||||||||||||||||
Db        274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy       1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy       1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
              |||||||||||||||   |||||||||||||:::|||   ||||||   ||||||||||||
Db        314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332

Qy       1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
              |||   |||         |||   ||||||||||||      :::||||||||||||
Db        333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy       1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
              ||||||   |||:::|||||||||||||||||||||      ||||||:::   ||||||
Db        353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371
```

```
Qy       1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db        372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391
Qy       1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db        392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411
Qy       1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db        412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431
Qy       1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db        432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451
Qy       1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db        452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471
Qy       1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db        472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491
Qy       1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db        492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511
Qy       1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db        512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531
Qy       1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db        532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551
Qy       1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db        552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571
Qy       1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db        572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591
Qy       2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db        592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611
Qy       2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db        612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631
Qy       2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db        632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651
Qy       2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db        652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
Qy       2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db        672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691
Qy       2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db        692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711
Qy       2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db        712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
Qy       2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db        732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751
Qy       2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db        752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
Qy       2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db        772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791
Qy       2658 CGCTGGGCCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
Db        792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811
Qy       2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
Db        812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831
Qy       2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
Db        832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851
Qy       2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
Db        852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871
Qy       2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
Db        872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891
Qy       2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
Db        892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
Qy       3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
Db        912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931
Qy       3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
Db        932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951
Qy       3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
Db        952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967
```

RESULT 8
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

```
Alignment Scores:
Pred. No.:                 0             Length:          968
Score:                     4332.50       Matches:         820
Percent Similarity:        89.86%        Conservative:    57
Best Local Similarity:     84.02%        Mismatches:      90
Query Match:               59.82%        Indels:          9
DB:                        4             Gaps:            5

US-09-302-812-1 (1-4070) x US-09-511-477-6 (1-968)

Qy    258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
Db      1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19

Qy    318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
Db     20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy    378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
Db     40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy    438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
Db     60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy    498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
Db     80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy    558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
Db     97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy    618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
Db    117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy    678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
Db    137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy    738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db    157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy    798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
Db    177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy    858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
Db    197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy    918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db    217 LysProThrValGlyAspGlyGln---------AlaArgSerAsnCysLysCysSerGly 233

Qy    978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db    234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy   1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db    254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy   1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db    274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy   1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db    294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy   1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db    314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332

Qy   1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db    333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy   1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db    353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy   1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db    372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy   1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db    392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy   1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db    412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy   1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db    432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy   1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db    452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy   1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db    472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy   1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db    492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy   1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db    512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy   1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db    532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy   1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db    552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy   1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db    572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy   2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db    592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy   2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db    612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy   2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db    632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy   2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db    652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
```

```
Qy      2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db       672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy      2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db       692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy      2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db       712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy      2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db       732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy      2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db       752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy      2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db       772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy      2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
Db       792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy      2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
Db       812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy      2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
Db       832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy      2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
Db       852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy      2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
Db       872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy      2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
Db       892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy      3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
Db       912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy      3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
Db       932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy      3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
Db       952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967
```

RESULT 9
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

Alignment Scores:

Pred. No.:	0	Length:	968
Score:	4332.50	Matches:	820
Percent Similarity:	89.86%	Conservative:	57
Best Local Similarity:	84.02%	Mismatches:	90
Query Match:	59.82%	Indels:	9
DB:	4	Gaps:	5

US-09-302-812-1 (1-4070) x US-09-511-507-6 (1-968)

```
Qy       258 ATGAGTGCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCCGCTGCA 317
Db         1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19

Qy       318 ACTTCTCCGCCGGCCGCCTCGGACGCCCGGAGCTTCCCCGGCAGGCAGAGGCGCGTCCTC 377
Db        20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy       378 GATTCCAAGGACGCTCCGGTGCAGTTCAGGGTCCCGCCGTCCTCGTCAGGCTGCGCCCTG 437
Db        40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy       438 GGCCGGGCGGGACAGCACCGAGGCAGCGCCACCTCTCTTGTTTTCAAACAGAAGACTATA 497
Db        60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy       498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTGCATAGTAAA 557
Db        80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy       558 GAAAACAACAATACAAGAGAAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAT 617
Db        97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy       618 CAACATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTTGATAAGTCACCAGTT 677
Db       117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy       678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAAGTGGCAGAAT 737
Db       137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy       738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db       157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy       798 GAGCAGTTCAGTAATGCTAATGTCGATCAGTCGTCCCCAAAGGATGATCACAGTGACACA 857
Db       177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy       858 AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCGAATGCA 917
Db       197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy       918 AAGCAGACGATGGAAGATGAACAGGGCAGAGAAGCCAGAAGCCACCAGAAGTGTGGCAAG 977
Db       217 LysProThrValGlyAspGlyGln---------AlaArgSerAsnCysLysCysSerGly 233

Qy       978 GCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAGGAGACAGACGTGGTGTCC 1037
Db       234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253
```

```
Qy  1038 GAGAGCCCCTTGTCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCCAAC 1097
Db   254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy  1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCTCCTCCATTTGAGAAAGAAAGTGAA 1157
Db   274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy  1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAAATAGTTGTCAGGATTCAGAAGCAGAT 1217
Db   294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy  1218 GAAGAGACAAGTCCAGGTTTTGATGAACAGGAAGATAGCAGTTCTGCTCAAACAGCAAAT 1277
Db   314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332

Qy  1278 AAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCCTCTGCT 1337
Db   333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy  1338 AAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGAGCTGGAATG 1397
Db   353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy  1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457
Db   372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy  1458 TCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGAGTGCCCAAA 1517
Db   392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy  1518 GCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAAAGGAAGATC 1577
Db   412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy  1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAG 1637
Db   432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy  1638 GAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAAT 1697
Db   452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy  1698 CACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTTCCTAAACCTTTCCCA 1757
Db   472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy  1758 ACACATTTTAAAGATTTGTGGGACAACAAGCATGTTAAGATGCCTTGTTCAGAACAAAAC 1817
Db   492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy  1818 TTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGGCAGCCGGTGGGAACTCATT 1877
Db   512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy  1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db   532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy  1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997
Db   552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy  1998 GTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGGTGAAA 2057
Db   572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy  2058 ATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATG 2117
Db   592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy  2118 AATCATTCCATCACAATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177
Db   612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy  2178 TGCACGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATT 2237
Db   632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy  2238 AACTTCAATCGGTTGTTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTC 2297
Db   652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy  2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAAGA 2357
Db   672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy  2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCAT 2417
Db   692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy  2418 GTCACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCA 2477
Db   712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy  2478 AACCGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2537
Db   732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy  2538 TTAATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db   752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy  2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657
Db   772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy  2658 CGCTGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACT 2717
Db   792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy  2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAG 2777
Db   812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy  2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGTCCTGGAGTTTCTTCA 2837
Db   832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy  2838 GAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897
Db   852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy  2898 AGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTTAT 2957
Db   872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy  2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017
Db   892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy  3018 GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACAATGAAGAA 3077
Db   912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy  3078 TGCAGAAACTGCTCCACCCCCGGACCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
Db   932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy  3138 GTTGAGTCCTGTACACAGACCACCAACCAGCCGGGACAAAGGACGGGG 3185
Db   952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967
```

RESULT 10

```
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-302-812-8

Alignment Scores:
Pred. No.:              2.46e-93      Length:         768
Score:                  1063.50       Matches:        244
Percent Similarity:     57.31%        Conservative:   93
Best Local Similarity:  41.50%        Mismatches:     198
Query Match:            14.69%        Indels:         53
DB:                     4             Gaps:           15

US-09-302-812-1 (1-4070) x US-09-302-812-8 (1-768)

Qy       1506 AGAGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACA 1565
               |||:::   |||:::  |||               :::   |||   :::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1566 GAAAGGAAGATCCCTAAATACATTCCACCTCACCTT------TCTCCAGATAAGAAATGG 1619
                                                |||   :::|||   |||   |||   |||
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1620 CTTGGAACTCCTATTGAGGAGATG------AGGAGAATGCCAAGGTGTGGGATCCGGCTG 1673
                 |||      :::|||   :::               |||   |||            |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATA 1733
              ||||||:::                  |||   |||         :::            |||   |||
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln---------LeuProIle 111

Qy       1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTTAAAGATTTGTGGGACAACAAGCATGTT 1793
                 |||      |||   |||                     ||||||:::::::|||||||
Db        112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130

Qy       1794 AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCT 1853
              ::::::||||||:::   ::::::      ||||||   |||:::   :::|||
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1854 GCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAG 1913
                       |||||||||:::|||:::   |||||||||         :::         :::
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1914 AACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1973
                 |||:::   ||||||:::   ||||||      |||      :::|||   |||
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1974 GCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAG 2033
              ||||||            |||:::   |||:::|||:::|||      :::        ::::::
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       2034 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCA 2093
                 :::||||||   ::::::::::::||||||   ||||||::::::            |||
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2094 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCACAGGAACAGATTGCC 2153
              :::||||||||||||            |||   |||:::::::::||||||:::||||||:::
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2154 AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA 2210
                 ||||||||||||||||||   ||||||||||||||||||||||||   |||   |||
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2211 TCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCA 2270
              ||||||||||||::::::||||||||||||||||||||||||||||::::::
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2271 AGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2324
                       ||||||||||||   :::         |||||||||||||||         ||||||
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2325 AAAAAA------------CCCACTGGGTTGGTGACATTCACAAGACAGAGT------CTT 2366
              :::                  |||||||||:::|||||||||      |||:::|||
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTG------ACTCGACTGCATGTC 2420
              |||            :::|||   :::            |||                  |||||||||
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2421 ACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
                 ||||||||||||||         |||      |||:::||||||||||||||||||||||
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
              :::::::::||||||||||||               |||      |||||||||||||||||||:::
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
              |||   ||||||||:::|||      :::||||||||||||         |||      |||
Db        391 IleCysProGluLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
              |||::::::   |||   |||:::||||||      ||||||||||||||||      ::::::
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2661 TGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACTGAG 2720
              |||:::   :::   ||||||   :::   ||||||            ||||||   |||
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy       2721 ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAG 2780
              ||||||||||||||||||||||||||      :::         |||:::         :::
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2781 ATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGT------------------ 2822
              :::   |||||||||||||||||||||   ||||||:::
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2823 CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCC 2882
              ||||||                        |||||||||||||||||||||||||||
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2883 TTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG 2942
              ||||||||||||:::   ||||||||||||:::|||::::::         |||
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2943 CGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002
              |||   :::   |||:::|||||||||:::   |||            |||   :::   |||
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy       3003 CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGA 3062
                                   ||||||   :::::::::         :::||||||
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       3063 TATTACAATGAAGAATGCAGAAACTGCTCCACCCCCGGACCA------------------ 3104
                 |||:::         :::   |||:::   |||
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
```

```
Qy       3105 GACATCAAGCTTTATCCATTCATATACCATGCAGTTGAGTCCTGTACACAGACCACCAAC 3164
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3165 CAGCCGGGACAAAGGACGGGGGCC 3188
Db        599 ValProGlyGluGlyAlaSerAla 606

RESULT 11
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-511-477-8

Alignment Scores:
Pred. No.:                2.46e-93       Length:         768
Score:                    1063.50        Matches:        244
Percent Similarity:       57.31%         Conservative:   93
Best Local Similarity:    41.50%         Mismatches:     198
Query Match:              14.69%         Indels:         53
DB:                       4              Gaps:           15

US-09-302-812-1 (1-4070) x US-09-511-477-8 (1-768)

Qy       1506 AGAGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACA 1565
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1566 GAAAGGAAGATCCCTAAATACATTCCACCTCACCTT------TCTCCAGATAAGAAATGG 1619
Db         63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1620 CTTGGAACTCCTATTGAGGAGATG------AGGAGAATGCCAAGGTGTGGGATCCGGCTG 1673
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATA 1733
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln---------LeuProIle 111

Qy       1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTAAAGATTTGTGGACAACAAGCATGTT 1793
Db        112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130

Qy       1794 AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCT 1853
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1854 GCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAG 1913
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1914 AACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1973
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1974 GCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAG 2033
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       2034 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCA 2093
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2094 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCACAGGAACAGATTGCC 2153
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2154 AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA 2210
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2211 TCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCA 2270
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2271 AGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2324
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2325 AAAAAA------------CCCACTGGGTTGGTGACATTCACAAGACAGAGT------CTT 2366
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTG------ACTCGACTGCATGTC 2420
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2421 ACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
Db        391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2661 TGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACTGAG 2720
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy       2721 ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAG 2780
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2781 ATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGT------------------ 2822
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2823 CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCC 2882
Db        491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy       2883 TTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG 2942
Db        503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy       2943 CGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002
Db        523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542
```

```
Qy       3003 CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGA 3062
                                    ||||||  ::::::::          :::||||||
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       3063 TATTACAATGAAGAATGCAGAAACTGCTCCACCCCCGGACCA------------------ 3104
                 |||:::         :::     |||:::     |||
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy       3105 GACATCAAGCTTTATCCATTCATATACCATGCAGTTGAGTCCTGTACACAGACCACCAAC 3164
                       |||||||||   ||||||           |||          :::     :::
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3165 CAGCCGGGACAAAGGACGGGGGCC 3188
                 ||||||:::         |||
Db        599 ValProGlyGluGlyAlaSerAla 606
```

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Alignment Scores:

Pred. No.:	2.46e-93	Length:	768
Score:	1063.50	Matches:	244
Percent Similarity:	57.31%	Conservative:	93
Best Local Similarity:	41.50%	Mismatches:	198
Query Match:	14.69%	Indels:	53
DB:	4	Gaps:	15

US-09-302-812-1 (1-4070) x US-09-511-507-8 (1-768)

```
Qy       1506 AGAGTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACA 1565
               |||:::   |||:::   |||              :::   |||   :::   :::
Db         45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy       1566 GAAAGGAAGATCCCTAAATACATTCCACCTCACCTT------TCTCCAGATAAGAAATGG 1619
                                               |||   :::|||      |||   |||      |||
Db         63 -----------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy       1620 CTTGGAACTCCTATTGAGGAGATG------AGGAGAATGCCAAGGTGTGGGATCCGGCTG 1673
                 |||      :::|||   :::            |||   |||               |||
Db         75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy       1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATA 1733
               ||||||:::                     |||      :::         |||   |||
Db         95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln---------LeuProIle 111

Qy       1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTAAAGATTTGTGGGACAACAAGCATGTT 1793
                 |||   |||   |||   |||                  ||||||::::::||||||
Db        112 ArgGluThrPro---ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130

Qy       1794 AAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGAGCT 1853
               ::::::|||||||:::   ::::::   ||||||   |||:::   :::|||
Db        131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy       1854 GCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAG 1913
                  |||||||||:::|||:::   |||||||||         :::            :::
Db        151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy       1914 AACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1973
                 |||:::   ||||||:::   ||||||      |||         :::|||   |||
Db        171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy       1974 GCTTTGATTGATTTCTGGGATAAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAG 2033
              ||||||               |||:::   |||:::|||:::|||   :::      ::::::
Db        191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy       2034 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGTACCCAGCCA 2093
                 :::||||||   ::::::::::::||||||   ||||||::::::         |||
Db        211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy       2094 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCACAGGAACAGATTGCC 2153
              :::||||||||||||                |||   |||::::::::::||||||:::||||||:::
Db        231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy       2154 AGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGACGCAATGCC---AAGATGAAA 2210
                 ||||||||||||||||||   ||||||||||||||||||         |||   |||
Db        251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy       2211 TCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTGTTTGAAGGACGTTCATCA 2270
              ||||||||||||::::::||||||||||||||||||||||||||||||::::::
Db        271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy       2271 AGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2324
                       ||||||||||||   :::      ||||||||||||||||      ||||||
Db        291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy       2325 AAAAAA------------CCCACTGGGTTGGTGACATTCACAAGACAGAGT------CTT 2366
              :::                     |||||||||:::|||||||||      |||:::|||
Db        311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy       2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTG------ACTCGACTGCATGTC 2420
              |||         :::|||   :::                |||         |||||||||
Db        331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy       2421 ACTTACGAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
                    |||||||||||||||         |||   |||:::|||||||||||||||||||||
Db        351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy       2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
              ::::::::::|||||||||||||         |||   |||||||||||||||||||||:::
Db        371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy       2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
              |||   |||||||||:::|||      :::||||||||||||      |||            |||
Db        391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy       2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
              |||::::::   |||      |||:::||||||      |||||||||||||||      ::::::
Db        411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy       2661 TGGGCCCGGAGCCATGAAGACAGGAGCGAAAGGGACGACTGGCAGAGGCGCACGACTGAG 2720
              |||:::   :::      ||||||   :::      ||||||            ||||||   |||
Db        431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy       2721 ATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCAGTTTGTGCCCGAGAAG 2780
              ||||||||||||||||||||||||||||         :::      |||:::   :::
Db        451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy       2781 ATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTTCTTCGT------------------ 2822
              :::   |||||||||||||||||||||||||||   ||||||:::
Db        471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy       2823 CCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAACTGGGGCTGTGGTGCC 2882
              ||||||                              |||||||||||||||||||||||||||
```

```
Db      491 ProGly-----------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy     2883 TTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCAGCTGCTGTAGCTGAG 2942
            ||||||||||||:::  |||||||||||:::|||::::::          |||
Db      503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy     2943 CGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002
            |||   :::   |||:::|||||||||:::   |||        |||   :::   |||
Db      523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542

Qy     3003 CATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGA 3062
                                    ||||||   :::::::::      :::||||||
Db      543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy     3063 TATTACAATGAAGAATGCAGAAACTGCTCCACCCCCGGACCA------------------ 3104
               |||:::          :::      |||:::      |||
Db      562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy     3105 GACATCAAGCTTTATCCATTCATATACCATGCAGTTGAGTCCTGTACACAGACCACCAAC 3164
                  |||||||||   ||||||           |||            :::     :::
Db      582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy     3165 CAGCCGGGACAAAGGACGGGGGCC 3188
               ||||||:::      |||
Db      599 ValProGlyGluGlyAlaSerAla 606
```

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-302-812-10

Alignment Scores:

Pred. No.:	1.37e-35	Length:	726
Score:	465.00	Matches:	204
Percent Similarity:	38.19%	Conservative:	116
Best Local Similarity:	24.34%	Mismatches:	282
Query Match:	6.42%	Indels:	236
DB:	4	Gaps:	35

US-09-302-812-1 (1-4070) x US-09-302-812-10 (1-726)

```
Qy      852 GACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCG 911
            |||      :::::::::   :::|||   :::   ::::::         |||      |||
Db       10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28

Qy      912 AATGCAAAGCAGACGATGGAA------------GATGAACAGGGCAGAGAAGCCAGAAGC 959
            :::         |||||||:::               |||   |||            |||
Db       29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48

Qy      960 CACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAG 1019
               :::                       |||   |||                     :::
Db       49 LysGlu-----------------AspProGluGlu------------------ProLys 56

Qy     1020 GAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGAT--------- 1070
               ||||||         |||      ||||||         ||||||
Db       57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76

Qy     1071 --------------------GTTGGTACTGGACTGAAAAATGCC--------------- 1094
                                |||            :::|||
Db       77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96

Qy     1095 -----------------------AACAGATTGAATAGACAAGAAAGTAGTCTAGGAAAT 1130
                                   |||         |||   |||      :::|||   |||
Db       97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116

Qy     1131 ---------TCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTA--- 1178
                  :::   |||   |||::::::            :::|||   :::||||||
Db      117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136

Qy     1179 --------------------------------GATAATTCCAAAAATAGTTGTCAGGAT 1205
                                            :::   :::         :::|||
Db      137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156

Qy     1206 TCAGAAGCAGATGAAGAGACAAGTCCAGGTTTT------GATGAACAGGAA--------- 1250
               |||      ::::::   |||               |||:::||||||
Db      157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176

Qy     1251 ------------------------------------------GATAGCAGTTCTGCTCAA 1268
                                                      |||         :::
Db      177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196

Qy     1269 ACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGG 1328
            ||||||   |||         :::      |||||||||:::
Db      197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer--------------- 211

Qy     1329 TCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGA 1388
                                                      |||   |||   :::|||:::
Db      212 ------------------------------------ThrPheValGlyGluAspSerLys 219

Qy     1389 GCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448
            |||                  |||      |||:::|||   |||:::
Db      220 Ala------------ThrLysThrValArgThrSerSerSerSerPheLeuSerThr--- 234

Qy     1449 TGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGA 1508

Db      234 ------------------------------------------------------------ 234

Qy     1509 GTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAA 1568
            |||         |||         |||   :::         |||         :::   |||
Db      235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254

Qy     1569 AGGAAGATCCCTAAATACATTCCACCTCACCTTTCT------CCAGATAAGAAATGGCTT 1622
            :::   :::      :::         :::|||:::      ||||||
Db      255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp------------ 270

Qy     1623 GGAACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTG 1682
                                                         |||      :::
Db      271 ------------------------------------------------LeuAsnLysVal 274

Qy     1683 AGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTT 1742
               |||      |||:::                          |||      :::
Db      275 AspProAspArgAsnTyr---------------------------ArgTyrCysThrIle 285

Qy     1743 CCTAAACCTTTCCCA---ACACATTTAAAGATTTGTGGGACAACAAGCAT--------- 1790
            |||      ||||||      :::      |||      ||||||::::::
Db      286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304

Qy     1791 GTTAAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1850
            :::   :::|||          :::                  |||      |||   |||
Db      305 IleValLeuProGlnArgTrpArgGluPhe------------AspSerArgGlyArgArg 320

Qy     1851 GCTGCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCC 19

Db      320 ------------------------------------------------------------

Qy     1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT---------GTGGCATATTCTAAGA
```

```
                   :::|||:::          :::                    |||   |||
Db     321 ---------ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337

Qy    1962 TGGGACTTTACAGCT--------------------TTGATTGATTTCTGGGAT------ 1994
            :::      |||                        |||    :::    |||:::
Db     338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAsp 357

Qy    1995 ---AAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051
                 :::      :::     |||      :::|||       :::    ::::::
Db     358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377

Qy    2052 GTG-------------------------------------AAAATTGCACTCTGTCTG 2072
            |||                                      :::||||||
Db     378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397

Qy    2073 CCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACA 2132
              :::|||      :::    |||      |||          :::        |||   |||
Db     398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416

Qy    2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192
            :::|||    :::|||    |||:::|||:::|||           ||||||
Db     417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe--------------- 431

Qy    2193 CGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTG 2252
                 |||:::     |||     :::|||                   |||    |||:::
Db     432 ------AlaArgProAspSerProPheSer-----------------PheCysArgIle 443

Qy    2253 TTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGA 2312
                          |||         |||||||||||     ||||||    ||||||
Db     444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463

Qy    2313 AGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTC------ACAAGACAGAGTCTT 2366
            :::::::::          |||    |||    |||:::|||        |||:::
Db     464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483

Qy    2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTAC 2426
            :::    |||    |||       :::     :::||||||    :::          |||    :::
Db     484 AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503

Qy    2427 ------GAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
                  |||         ||||||                      ||||||||||||||||||
Db     504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522

Qy    2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
                 :::||||||||||||    :::     |||     ||||||||||||||||||||||||
Db     523 GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542

Qy    2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
            :::    ||||||::::::|||          |||       |||    :::         |||
Db     543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562

Qy    2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
            :::    |||     |||          :::|||     |||||||||||     |||    :::
Db     563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582

Qy    2661 TGGCCC---------------CGGAGCCATGAAGACAGGAGCGAA---AGGGACGACTGG 2702
            ||||||                      |||    ::::::    :::|||     ||||||    :::
Db     583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602

Qy    2703 CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC--------- 2753
                 |||          |||     :::||||||||||||:::     |||:::
Db     603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622

Qy    2754 ------TACCTCGACCAGTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAGGCTTAC 2807
                          :::|||                |||    ||||||:::      ||||||
Db     623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642

Qy    2808 TGTGGATTTCTTCGTCCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAAC 2867
               ||||||:::         |||                |||:::          :::   ||||||
Db     643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp 661

Qy    2868 TGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCA 2927
            ||||||||||||||||||        ||||||           ||||||      :::||||||::::::|||
Db     662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681

Qy    2928 GCTGCTGTAGCTGAGCGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGA 2987
            |||      ||||||:::|||      :::     :::     :::||||||:::      ||||||
Db     682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701

Qy    2988 GACATTTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAA 3041
                        :::          :::     ::::::     :::|||:::|||:::
Db     702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719
```

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-477-10

```
Alignment Scores:
Pred. No.:              1.37e-35        Length:          726
Score:                  465.00          Matches:         204
Percent Similarity:     38.19%          Conservative:    116
Best Local Similarity:  24.34%          Mismatches:      282
Query Match:            6.42%           Indels:          236
DB:                     4               Gaps:            35

US-09-302-812-1 (1-4070) x US-09-511-477-10 (1-726)

Qy     852 GACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTTGCG 911
            |||      :::::::::     :::|||    :::     ::::::          |||    |||
Db      10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28

Qy     912 AATGCAAAGCAGACGATGGAA------------GATGAACAGGGCAGAGAAGCCAGAAGC 959
            :::          ||||||:::                |||    |||          |||
Db      29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48

Qy     960 CACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAG 1019
               :::                 |||   |||                        :::
Db      49 LysGlu------------------AspProGluGlu------------------ProLys 56

Qy    1020 GAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGAT--------- 1070
                  ||||||            |||      ||||||         ||||||
Db      57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76

Qy    1071 ---------------------GTTGGTACTGGACTGAAAAATGCC--------------- 1094
                                |||          :::|||
Db      77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96

Qy    1095 ------------------------AACAGATTGAATAGACAAGAAAGTAGTCTAGGAAAT 1130
                                   |||          |||     |||     :::|||    |||
Db      97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116

Qy    1131 ---------TCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTA--- 1178
                    :::     |||     |||::::::          :::|||     :::||||||
Db     117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136
```

```
Qy      1179 --------------------------------GATAATTCCAAAAATAGTTGTCAGGAT 1205
                                              :::   :::         :::|||
Db       137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156

Qy      1206 TCAGAAGCAGATGAAGAGACAAGTCCAGGTTTT------GATGAACAGGAA--------- 1250
                |||      ::::::    |||            |||:::||||||
Db       157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176

Qy      1251 -----------------------------------------GATAGCAGTTCTGCTCAA 1268
                                                         |||       :::
Db       177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196

Qy      1269 ACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGG 1328
             ||||||    |||             :::        |||||||||:::
Db       197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer--------------- 211

Qy      1329 TCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGA 1388
                                                |||   |||    :::|||:::
Db       212 ------------------------------------ThrPheValGlyGluAspSerLys 219

Qy      1389 GCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448
             |||                        |||      |||:::|||   |||:::
Db       220 Ala-----------ThrLysThrValArgThrSerSerSerSerPheLeuSerThr--- 234

Qy      1449 TGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGA 1508

Db       234 ------------------------------------------------------------ 234

Qy      1509 GTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAA 1568
             |||         |||         |||    :::      |||       :::      |||
Db       235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254

Qy      1569 AGGAAGATCCCTAAATACATTCCACCTCACCTTTCT------CCAGATAAGAAATGGCTT 1622
             :::   :::       :::         :::|||:::          ||||||
Db       255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp------------ 270

Qy      1623 GGAACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTG 1682
                                                             |||      :::
Db       271 ------------------------------------------------LeuAsnLysVal 274

Qy      1683 AGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTT 1742
                |||      |||:::                             |||        :::
Db       275 AspProAspArgAsnTyr---------------------------ArgTyrCysThrIle 285

Qy      1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTTGTGGGACAACAAGCAT--------- 1790
             |||      ||||||   :::      |||          ||||||::::::
Db       286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304

Qy      1791 GTTAAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1850
             :::   :::|||      :::                  |||      |||     |||
Db       305 IleValLeuProGlnArgTrpArgGluPhe------------AspSerArgGlyArgArg 320

Qy      1851 GCTGCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCC 1910

Db       320 ------------------------------------------------------------ 320

Qy      1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT---------GTGGCATATTCTAAGAAA 1961
                      :::|||:::           :::               |||   |||
Db       321 ---------ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337

Qy      1962 TGGGACTTTACAGCT--------------------TTGATTGATTTCTGGGAT------ 1994
             :::      |||                       |||   :::   |||:::
Db       338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAsp 357

Qy      1995 ---AAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051
                :::      :::        |||      :::|||          :::   ::::::
Db       358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377

Qy      2052 GTG-------------------------------------AAAATTGCACTCTGTCTG 2072
             |||                                      :::||||||
Db       378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397

Qy      2073 CCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACA 2132
                :::|||      :::   |||      |||         :::      |||   |||
Db       398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416

Qy      2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192
             :::|||      :::|||      |||:::|||:::|||          ||||||
Db       417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe--------------- 431

Qy      2193 CGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTG 2252
                   |||:::      |||   :::|||                    |||   |||:::
Db       432 ------AlaArgProAspSerProPheSer------------------PheCysArgIle 443

Qy      2253 TTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGA 2312
                         |||         |||||||||||   ||||||   ||||||
Db       444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463

Qy      2313 AGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTC------ACAAGACAGAGTCTT 2366
             :::::::::            |||   |||   |||:::|||      |||:::
Db       464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483

Qy      2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTAC 2426
             :::   |||   |||      :::      :::||||||          |||   :::
Db       484 AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503

Qy      2427 ------GAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
                   |||      ||||||                     ||||||||||||||||||
Db       504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522

Qy      2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
                   :::||||||||||||   :::      |||   ||||||||||||||||||||||||
Db       523 GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542

Qy      2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
             :::   ||||||:::::::|||      |||      |||   :::
Db       543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562

Qy      2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
             :::   |||   |||      :::|||   ||||||||||||      |||   :::
Db       563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582

Qy      2661 TGGGCC--------------CGGAGCCATGAAGACAGGAGCGAA---AGGGACGACTGG 2702
             ||||||               |||   ::::::   :::|||   ||||||   :::
Db       583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602

Qy      2703 CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC--------- 2753
                |||         |||   :::||||||||||||:::   |||:::
Db       603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622

Qy      2754 ------TACCTCGACCAGTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAGGCTTAC 2807
                      :::|||               |||   ||||||:::   ||||||
Db       623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642

Qy      2808 TGTGGATTTCTTCGTCCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAAC 2867
                ||||||:::         |||            |||:::      :::   ||||||
Db       643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp 661

Qy      2868 TGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCA 2927
             ||||||||||||||||||      ||||||      ||||||   :::||||||::::::|||
Db       662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681

Qy      2928 GCTGCTGTAGCTGAGCGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGA 2987
             |||   ||||||:::|||   :::   :::   :::||||||:::   ||||||
Db       682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701

Qy      2988 GACATTTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAA 3041
                   :::         :::   ::::::   :::|||:::|||:::
Db       702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719
```

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.

```
; APPLICANT:  JACOBSON, Elaine L.
; APPLICANT:  AME, Jean-Christophe
; APPLICANT:  LIN, Winston
; TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION:  THEREWITH
; FILE REFERENCE:  NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE:  2000-02-23
; PRIOR APPLICATION NUMBER:  09/302,812
; PRIOR FILING DATE:  1999-04-30
; NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 10
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE:
US-09-511-507-10

Alignment Scores:
Pred. No.:              1.37e-35      Length:         726
Score:                  465.00        Matches:        204
Percent Similarity:     38.19%        Conservative:   116
Best Local Similarity:  24.34%        Mismatches:     282
Query Match:            6.42%         Indels:         236
DB:                     4             Gaps:           35

US-09-302-812-1 (1-4070) x US-09-511-507-10 (1-726)

Qy     852 GACACAAATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTGACACATGTAAAGCTTGCG 911
Db      10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28

Qy     912 AATGCAAAGCAGACGATGGAA-----------GATGAACAGGGCAGAGAAGCCAGAAGC 959
Db      29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48

Qy     960 CACCAGAAGTGTGGCAAGGCTTGCCATCCTGCAGAAGCCTGTGCAGGGTGTCAGCAGGAG 1019
Db      49 LysGlu-----------------AspProGluGlu-----------------ProLys 56

Qy    1020 GAGACAGACGTGGTGTCCGAGAGCCCCTTGTCGGACACTGGCTCTGAGGAT--------- 1070
Db      57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76

Qy    1071 --------------------GTTGGTACTGGACTGAAAAATGCC--------------- 1094
Db      77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96

Qy    1095 -----------------------AACAGATTGAATAGACAAGAAAGTAGTCTAGGAAAT 1130
Db      97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116

Qy    1131 ---------TCTCCTCCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGGATGTA--- 1178
Db     117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136

Qy    1179 -------------------------------GATAATTCCAAAAATAGTTGTCAGGAT 1205
Db     137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156

Qy    1206 TCAGAAGCAGATGAAGAGACAAGTCCAGGTTTT------GATGAACAGGAA--------- 1250
Db     157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176

Qy    1251 ----------------------------------------GATAGCAGTTCTGCTCAA 1268
Db     177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196

Qy    1269 ACAGCAAATAAACCTTCAAGGTTCCAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGG 1328
Db     197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer--------------- 211

Qy    1329 TCCTCTGCTAAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGA 1388
Db     212 -----------------------------------ThrPheValGlyGluAspSerLys 219

Qy    1389 GCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448
Db     220 Ala------------ThrLysThrValArgThrSerSerSerSerPheLeuSerThr--- 234

Qy    1449 TGCAGAAATTCTAAGCAACATGGGAGAAAGGATTCTAAAATCACAGATCATTTCATGAGA 1508
Db     234 ------------------------------------------------------------ 234

Qy    1509 GTGCCCAAAGCAGAGGACAAAAGAAAAGAACAATGTGAAATGAAACATCAAAGAACAGAA 1568
Db     235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254

Qy    1569 AGGAAGATCCCTAAATACATTCCACCTCACCTTTCT------CCAGATAAGAAATGGCTT 1622
Db     255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----------- 270

Qy    1623 GGAACTCCTATTGAGGAGATGAGGAGAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTG 1682
Db     271 ------------------------------------------------LeuAsnLysVal 274

Qy    1683 AGACCATCTGCCAATCACACAGTGACTATTCGGGTAGATCTTTTGCGAATAGGAGAAGTT 1742
Db     275 AspProAspArgAsnTyr--------------------------ArgTyrCysThrIle 285

Qy    1743 CCTAAACCTTTCCCA---ACACATTTAAAGATTTGTGGGACAACAAGCAT--------- 1790
Db     286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304

Qy    1791 GTTAAGATGCCTTGTTCAGAACAAAACTTGTACCCTGTGGAAGATGAGAATGGTGAGCGA 1850
Db     305 IleValLeuProGlnArgTrpArgGluPhe------------AspSerArgGlyArgArg 320

Qy    1851 GCTGCAGGCAGCCGGTGGGAACTCATTCAGACTGCACTTCTCAACAGGCTCACTCGGCCC 1910
Db     320 ------------------------------------------------------------ 320

Qy    1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT---------GTGGCATATTCTAAGAAA 1961
Db     321 ---------ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337

Qy    1962 TGGGACTTTACAGCT--------------------TTGATTGATTTCTGGGAT------ 1994
Db     338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAsp 357

Qy    1995 ---AAGGTACTAGAAGAAGCAGAAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATG 2051
Db     358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377

Qy    2052 GTG---------------------------------------AAAATTGCACTCTGTCTG 2072
Db     378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397

Qy    2073 CCAAATATTTGTACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACA 2132
Db     398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416

Qy    2133 ATGTCACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACGTTTCCACGA 2192
Db     417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe--------------- 431

Qy    2193 CGCAATGCCAAGATGAAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGGTTG 2252
Db     432 ------AlaArgProAspSerProPheSer-----------------PheCysArgIle 443

Qy    2253 TTTGAAGGACGTTCATCAAGGAAACCAGAGAAGCTTAAAACGCTCTTCTGCTACTTTAGA 2312
Db     444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463

Qy    2313 AGAGTCACAGAGAAAAAACCCACTGGGTTGGTGACATTC------ACAAGACAGAGTCTT 2366
```

```
                 :::::::::          |||    |||    |||:::|||      |||:::
Db        464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483

Qy       2367 GAAGATTTTCCAGAGTGGAAAGATGTGAAAAACTCCTGACTCGACTGCATGTCACTTAC 2426
              :::    |||   |||        :::    :::||||||    :::          |||    :::
Db        484 AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503

Qy       2427 ------GAAGGTACCATAGAAGGAAACGGCCAGGGCATGCTACAGGTGGATTTTGCAAAC 2480
                    |||      ||||||                         ||||||||||||||||
Db        504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522

Qy       2481 CGTTTCGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2540
                    :::|||||||||||    :::   |||   |||||||||||||||||||||
Db        523 GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542

Qy       2541 ATCAACCCTGAGTTGATTGTTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2600
              :::    ||||||::::::|||       |||      |||     :::          |||
Db        543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562

Qy       2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660
              :::    |||   |||       :::|||   ||||||||||||      |||     :::
Db        563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582

Qy       2661 TGGGCC---------------CGGAGCCATGAAGACAGGAGCGAA---AGGGACGACTGG 2702
              ||||||                   |||    ::::::    :::|||   ||||||     :::
Db        583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602

Qy       2703 CAGAGGCGCACGACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGACGC--------- 2753
                 |||          |||    :::||||||||||||:::    |||:::
Db        603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622

Qy       2754 ------TACCTCGACCAGTTTGTGCCCGAGAAGATCAGACGGGAGCTTAACAAGGCTTAC 2807
                          :::|||                  |||    ||||||:::    ||||||
Db        623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642

Qy       2808 TGTGGATTTCTTCGTCCTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAAC 2867
                 ||||||:::        |||               |||:::       :::    ||||||
Db        643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp 661

Qy       2868 TGGGGCTGTGGTGCCTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCCTGGCA 2927
              |||||||||||||||||    ||||||         ||||||    :::||||||::::::|||
Db        662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681

Qy       2928 GCTGCTGTAGCTGAGCGAGACGTGGTTTATTTCACCTTTGGGGACTCAGAACTGATGAGA 2987
              |||    ||||||:::|||    :::    :::    :::||||||:::    ||||||
Db        682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701

Qy       2988 GACATTTACAGCATGCATACATTCCTCACTGAGAGGAAACTGACTGTTGGAGAA 3041
                       :::           :::    ::::::    :::|||:::|||:::
Db        702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719
```

Search completed: May 26, 2004, 17:17:29
Job time : 168.05 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 26, 2004, 15:32:43 ; Search time 34.0417 Seconds
                                           (without alignments)
                                           12447.891 Million cell updates/sec

Title:          US-09-302-812-3
Perfect score:  7208
Sequence:       1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150053_5664/app_query.fasta_1.12437
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_128_@runat_26052004_150053_5664 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	186.5	2.6	2476	1	ATRX_MOUSE	Q61687 mus musculu
2	181.5	2.5	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
3	181.5	2.5	5085	1	PCLO_RAT	Q9jks6 rattus norv
4	174	2.4	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
5	172.5	2.4	5038	1	PCLO_MOUSE	Q9qyx7 mus musculu
6	162	2.2	2492	1	ATRX_PANTR	Q7yqm4 pan troglod
7	161	2.2	2492	1	ATRX_HUMAN	P46100 homo sapien
8	161	2.2	2492	1	ATRX_PONPY	Q7yqm3 pongo pygma
9	160	2.2	1957	1	SPOF_SCHPO	Q10411 schizosacch
10	159.5	2.2	1183	1	BMS1_YEAST	Q08965 saccharomyc
11	159	2.2	406	1	SR40_YEAST	P32583 saccharomyc
12	159	2.2	1031	1	YDG9_SCHPO	Q10496 schizosacch
13	159	2.2	1462	1	NKCR_HUMAN	P30414 homo sapien
14	159	2.2	2418	1	SPCA_HUMAN	P02549 homo sapien
15	158.5	2.2	6632	1	UN89_CAEEL	O01761 caenorhabdi
16	158	2.2	1311	1	ATRX_DROME	Q9gqn5 drosophila
17	155	2.2	1138	1	BM2K_MOUSE	Q91z96 mus musculu
18	154.5	2.1	646	1	SG1_BOVIN	P23389 bos taurus
19	154.5	2.1	3259	1	GOB1_HUMAN	Q14789 homo sapien
20	154	2.1	1972	1	MYHB_MOUSE	O08638 mus musculu
21	154	2.1	2238	1	GOA4_MOUSE	Q91vw5 mus musculu
22	153.5	2.1	1878	1	BRC1_CANFA	Q95153 canis famil
23	152.5	2.1	2415	1	SPCA_MOUSE	P08032 mus musculu
24	152	2.1	644	1	NFM_RABIT	P54938 oryctolagus
25	152	2.1	1447	1	BUD4_YEAST	P47136 saccharomyc
26	152	2.1	5560	1	SPEN_DROME	Q8sx83 drosophila
27	151.5	2.1	1338	1	ACIN_MOUSE	Q9jix8 mus musculu
28	150.5	2.1	1312	1	RA50_YEAST	P12753 saccharomyc
29	150.5	2.1	1972	1	MYHB_RABIT	P35748 oryctolagus
30	150	2.1	2137	1	SPCB_HUMAN	P11277 homo sapien
31	150	2.1	3969	1	HRX_HUMAN	Q03164 homo sapien
32	149.5	2.1	1210	1	AF4_HUMAN	P51825 homo sapien
33	149	2.1	1093	1	YKD5_CAEEL	Q03563 caenorhabdi
34	148.5	2.1	472	1	YWIE_CAEEL	Q23525 caenorhabdi
35	148.5	2.1	677	1	SG1_HUMAN	P05060 homo sapien
36	148.5	2.1	1928	1	MYS1_YEAST	P08964 saccharomyc
37	148.5	2.1	1969	1	MYSA_CAEEL	P12844 caenorhabdi
38	148	2.1	627	1	ABPX_YEAST	Q08641 saccharomyc
39	148	2.1	1679	1	GCC2_MOUSE	Q8chg3 mus musculu
40	148	2.1	3210	1	CENF_HUMAN	P49454 homo sapien
41	147	2.0	943	1	YM11_YEAST	P39523 saccharomyc
42	146.5	2.0	667	1	CYL1_BOVIN	P35662 bos taurus
43	146.5	2.0	1560	1	TENN_MOUSE	Q80z71 mus musculu
44	146.5	2.0	3214	1	BPA1_HUMAN	Q03001 homo sapien
45	146	2.0	5120	1	PCLO_CHICK	Q9pu36 gallus gall

ALIGNMENTS

```
RESULT 1
ATRX_MOUSE
ID   ATRX_MOUSE     STANDARD;      PRT;  2476 AA.
AC   Q61687;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked nuclear protein)
DE   (Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38
DE   protein).
GN   ATRX OR XNP OR HP1BP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98213653; PubMed=9545503;
RA   Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
RT   "Comparison of the human and murine ATRX gene identifies highly
RT   conserved, functionally important domains.";
RL   Mamm. Genome 9:400-403(1998).
RN   [2]
RP   SEQUENCE OF 325-1176 FROM N.A.
RX   MEDLINE=97133299; PubMed=8978696;
RA   le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA   Jeanmougin F., Losson R., Chambon P.;
RT   "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT   control of transcription by nuclear receptors.";
RL   EMBO J. 15:6701-6715(1996).
RN   [3]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
```

```
CC       gene expression by affecting chromatin.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF026032; AAC08741.1; -.
DR   EMBL; X99643; CAA67962.1; -.
DR   MGD; MGI:103067; Atrx.
DR   GO; GO:0000228; C:nuclear chromosome; IDA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING     219    267       PHD-TYPE.
FT   NP_BIND    1579   1586       ATP (POTENTIAL).
FT   SITE       1704   1707       DEGH BOX.
FT   DOMAIN      319    322       POLY-SER.
FT   DOMAIN      735    738       POLY-SER.
FT   DOMAIN     1001   1004       POLY-GLU.
FT   DOMAIN     1130   1135       POLY-SER.
FT   DOMAIN     1182   1185       POLY-SER.
FT   DOMAIN     1238   1245       POLY-ASP.
FT   DOMAIN     1484   1487       POLY-GLU.
FT   DOMAIN     1924   1931       POLY-SER.
FT   DOMAIN     2205   2208       POLY-LYS.
FT   DOMAIN     2245   2248       POLY-GLU.
FT   DOMAIN     2403   2408       POLY-GLN.
SQ   SEQUENCE   2476 AA;  278601 MW;  90A42B790FC4FF4C CRC64;

Alignment Scores:
Pred. No.:               0.000113      Length:          2476
Score:                   186.50        Matches:         216
Percent Similarity:      35.48%        Conservative:    142
Best Local Similarity:   21.41%        Mismatches:      379
Query Match:             2.59%         Indels:          273
DB:                      1             Gaps:            53

US-09-302-812-3 (1-4069) x ATRX_MOUSE (1-2476)

Qy         67 AGGCAGGCGCTGAGAAGGACGCGCAGTCCATCTCTCAGGTTAGTGAAATGAGGCTCTC 126
Db        967 ArgProGlyThrLysGlyLysLysAlaProAspLeuLysGly---------GluThrLeu 983

Qy        127 CGC-----------------GGGGCCGGCCCGGGGACAGTGCGCTGCTGGTCCCAGCAT 168
Db        984 LysArgGluGlnGluTrpAspSerSerSerAspGlyThrGluArgLeuProGluGluGlu 1003

Qy        169 GAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGC--------------------GAC 207
Db       1004 GluIleGlyProPheSerLysGlyIleLysGlnSerLysThrAspThrAlaGlyGlyGlu 1023

Qy        208 CCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAG 267
Db       1024 LysLysGlyLysLysTrpLysAspLysSerCysGluLysLysGluGluLeuSerAspSer 1043

Qy        268 GCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTC 327

Db       1044 ValAspLysLeuPro---GlyLysGlyAspSer--CysAspSerSerGluAspLysLysT 1062

Qy        328 GCCAG-------------------------------------CCTGCG---TCCCAGG 345
Db       1062 hrArgAsnArgValSerLeuArgGluLysLysArgPheSerLeuProAlaLysSerProG 1082

Qy        346 GCAGGCGG------------------------------------------------G 354
Db       1082 lyLysArgProGluCysSerSerSerAspThrGluLysSerLeuLysGlyGlnCysCysA 1102

Qy        355 ACAGCACAGAGGCAGCGCCACCTCGCTTGTT----TTCAAACAAAAGACTATTACCAGTT 410
Db       1102 spSerThrGluLysArgProLysArgIleAspLeuArgGluArgArgAsnSerSerSerL 1122

Qy        411 GGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGT------TTGGATAGTAAAG 464
Db       1122 ysArgAsnThrLysGluValLysSerAlaSerSerSerAspAlaGluGlySerSerG 1142

Qy        465 AAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACC 524
Db       1142 luAspAsnLysLysGlnLysLysGlnArgThrSerAlaLysLysLysThrGlyAsnThrL 1162

Qy        525 AACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTG 584
Db       1162 ysGluLysLysArgAsnSerLeuArgAlaThrProLysArgLysGlnValAspIleThrS 1182

Qy        585 AAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATG 644
Db       1182 erSerSerSerAspIleGlyAspAspAspGlnAsnSerAlaGlyGluGluSerSerAspG 1202

Qy        645 AAGGGAAA------------------------------------CACACG---------- 658
Db       1202 luGlnLysIleLysProValThrGluAsnLeuValLeuProSerHisThrGlyPheCysG 1222

Qy        659 -------------GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAG 704
Db       1222 lnSerSerGlyAspGluAlaLeuSerLysSerValProAlaThrVal------------ 1237

Qy        705 AGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAG 764
Db       1238 ----------------------------------------AspAspAspAspAspAspA 1244

Qy        765 ATAGTGAAGAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAA 824
Db       1244 snAspProGluAsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuS 1264

Qy        825 AGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGA 872
Db       1264 erSerAspGluAspGlySerSerAspAspGluProAspGlyGlyGlyLysLysArgIleG 1284

Qy        873 AGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAG 932
Db       1284 lyLysGlnSerGluGluSerProAlaAspAsp---GlyGluLeuArgArgGluGlnLeuA 1303

Qy        933 ACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAA 992
Db       1303 laValAsnGlnValAsnSerGluSerAspSerAspSerGluGlu-----------SerL 1319

Qy        993 AAAATGAC---------AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTC 1043
Db       1319 ysLysProArgTyrArgHisArgLeuLeuArgHisLysLeuThrLeuSerAspGlyGluS 1339

Qy       1044 CA---TTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATA 1100
Db       1339 erGlyGluGluLysProThrLysProLysGluHisLysGluAlaLysGlyArgAsnArgA 1359

Qy       1101 GTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATG 1160
Db       1359 rgLysValSerSerGluAspSerGluAspThr------------AspPheGlnGluSerG 1375

Qy       1161 GT---------------AGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAA 1205
```

```
Db   1375 lyValSerGluGluValSerGluSerGluAspGluGlnArgProArgThrArgSerAlaL 1395

Qy   1206 GAGACGCTGACATTGAATTTAGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGAT 1262

Db   1395 ysLysAlaGluLeuGluGluAsnGlnArgSerTyrLysGlnLysLysLysArgArgArgI 1415

Qy   1263 TACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAAC 1322

Db   1415 leLysValGln---GluAspSerSerSerGluAsnLysSerHisSerGluGluAspLysL 1434

Qy   1323 TACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGA---------------------A 1361

Db   1434 ysGluGlyAspGluGluAspGluGluAspGluAspGluAspGluGluAspGluAsnAspA 1454

Qy   1362 ATTCTAAGCAACATGGAAAA--------------------------AAGGATTCTAAAA 1394

Db   1454 spSerLysSerProGlyLysGlyArgLysLysIleArgLysIleLeuLysAspAspLysL 1474

Qy   1395 TC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGG 1451

Db   1474 euArgThrGluThrGlnAsnAlaLeuLysGluGluGluGluArgArgLysArgIleAlaG 1494

Qy   1452 AAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA---CCTCACCTTT 1508

Db   1494 luArgGluArgGluArg------GluLysLeuArgGluValIleGluIleGluAspAlaS 1512

Qy   1509 CTCCAGATAAGAAGTGGCTTGGAACTCCCATT---------------GAGGAGATGAGAA 1553

Db   1512 erProThrLysCysProIleThrThrLysLeuValLeuAspGluAsnGluGluThrLysG 1532

Qy   1554 GAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAA 1613

Db   1532 lu------------------------ProLeuValGlnValHisArgAsnMetValIleL 1544

Qy   1614 CTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATA 1673

Db   1544 ysLeu--------------------------------LysPro---------HisGlnV 1550

Qy   1674 AAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAA 1721

Db   1550 alAspGlyValGlnPheMetTrpAsp--------------------CysCysCysGluS 1563

Qy   1722 ATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCA 1781

Db   1563 erVal---------------GluLysThrLysLysSerProGlySerGlyCysIleLeuA 1578

Qy   1782 TTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGA 1841

Db   1578 laHisCysMetGlyLeuGlyLys---------------ThrLeuGlnValValSerPheL 1593

Qy   1842 AATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTGATCGATTTC---- 1894

Db   1593 euHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeuValValCysProL 1613

Qy   1895 -----------------TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATC 1937

Db   1613 euAsnThrAlaLeuAsnTrpMetAsnGluPheGluLys------------------TrpG 1627

Qy   1938 AGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGC 1997

Db   1627 lnGluGlyLeuAsnAspAsnGluLysLeuGlu------ValSerGluLeuAlaThrValL 1645

Qy   1998 CAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTG 2057

Db   1645 ysArgPro---------GlnGluArgSerTyrMetLeuGlnArgTrpGlnGluAspGlyG 1662

Qy   2058 CCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAAT 2117

Db   1662 lyValMetIleIleGly------------------------------------------- 1667

Qy   2118 CGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAA 2177

Db   1668 -----TyrGluMetTyr---------------ArgAsnLeuAlaGlnGlyArgAsnVal- 1680

Qy   2178 GGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAAC 2237

Db   1681 --LysSerArgLysLeuLysAspIlePhe------AsnLysAlaLeuValAspProGlyP 1698

Qy   2238 CTACTGGGTTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGAT 2297

Db   1698 roAspPheValValCysAspGluGlyHisIleLeuLysAsn------GluAlaSerAlaV 1716

Qy   2298 GTGAAAAAACCCTTG------------ACACGATTGCATGTCACTTACGAAGGTACC---- 2341

Db   1716 alSerLysAlaMetAsnSerIleLysSerArgArgArgIleIleLeuThrGlyThrProL 1736

Qy   2342 ------------------------------------------ATAGAAGAAAATGGCCAAG 2360

Db   1736 euGlnAsnAsnLeuIleGluTyrHisCysMetValAsnPheIleLysGluAsnLeuLeuG 1756

Qy   2361 GCATGCTACAGGTGGATTTTGCAAATCGTTTTGTT------------GGAGGTGGTGTAA 2408

Db   1756 lySerIleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGlyGlnCysA 1775

Qy   2409 CCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTT 2468

Db   1775 laAspSerThrMetVal------AspValArgValMetLysLysArgAlaHisIleLeuT 1793

Qy   2469 CACGGCTC-------------------------------------TTCACTGAGGTGCTG- 2491

Db   1793 yrGluMetLeuAlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLysPheLeuP 1813

Qy   2492 -----GATCACAATGAATGTCTAATTATCACAGGTACTGAG---CAGTACAGTGAATACA 2543

Db   1813 roProLysHisGluTyrValLeuAlaValArgMetThrAlaIleGlnCysLysLeuTyrG 1833

Qy   2544 CAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACG 2603

Db   1833 lnTyrTyr-LeuAspHisLeuThrGlyValGlyAsnSerThrGluGlyGlyArgGly--- 1851

Qy   2604 ACTGCGAGCGGCGCTGCACTG 2624

Db   1852 ---LysAlaGlyAlaLysLeu 1857
```

```
RESULT 2
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   PIR; T34036; T34036.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;
KW   DNA-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;

Alignment Scores:
Pred. No.:              0.000209      Length:         1359
Score:                  181.50        Matches:        182
Percent Similarity:     32.44%        Conservative:   131
Best Local Similarity:  18.86%        Mismatches:     328
Query Match:            2.52%         Indels:         324
DB:                     1             Gaps:           39

US-09-302-812-3 (1-4069) x ATRX_CAEEL (1-1359)

Qy      78 GAGAAGGACGCGCAGTCCATCTCTCTCAGGTTAGTGAAATGAGGCTCTCCGCGGGGCCGG 137
              ||||||  ||||||    :::       :::      |||  |||
Db      33 GluLysArgAlaGlnLysLeuLysGluLysArgGluArgGluGlyLysPro--------- 49

Qy     138 CCCGGGGACAGTGCGCTGCTGGTCCCAGCATGAATGCGGGCCCCGGCTGTGAACCCTGCA 197
                                                                   |||
Db      50 ---------------------------------------------------------ProPro 51

Qy     198 CCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCT 257
                |||||||||||
Db      52 LysLysArgProAla--------------------------------------------- 56

Qy     258 TTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCC 317

Db      56 ------------------------------------------------------------ 56

Qy     318 CACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCT 377
              :::   |||:::  :::|||           |||     |||           |||
Db      57 ---LysLysArgLysAlaSerSerSerGluGluAspAspAspAspGluGluGluSerPro 75

Qy     378 CGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAG 437
              |||     |||   |||                   :::   |||   ::::::  :
Db      76 ArgLysSerSerLysLys--------------------SerArgLysArgAlaLys--S 88

Qy     438 CGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGA 497
             ::|||||||||||     :::|||    |||                |||
Db      88 erGluSerGluSerAspGluSerAspGluGlu---------------GluAspArgLysL 103

Qy     498 GTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAA------AAATTAGTAAATG 551
               |||      |||             |||        |||        |||
Db     103 ysSerLysSerLysLysLysValAspGlnLysLysLysGluLysSerLysLysLysArgT 123

Qy     552 TTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGC 611
              :::   :::   |||   |||:::          |||:::       |||
Db     123 hrThrSerSerSerGluAspGluAspSerAspGluGluArgGluGln------LysSerL 141

Qy     612 ATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGG 671
              :::   :::            |||   |||            :::   :::|||    |
Db     141 ysLysLysSerLysLysThrLysLysGlnThrSerSerGluSerSerGlu---------G 158

Qy     672 AAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATC 731
             |||||||||   :::                  ::::::            ::::::   :
Db     158 luSerGluGluGluArgLysValLysLysSerLysLysAsnLysGluLysSerValLysL 178

Qy     732 GGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAAC 791
             ::        :::      :::   ||||||   |||   :::   ::::::   :::::::
Db     178 ysArgAlaGluThrSerGluGluSerAspGluAspGluLysProSerLysLysSerLysL 198

Qy     792 AGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCCA 851
             ::   |||           |||   |||   ::::::      :::|||   |||   :
Db     198 ysGlyLeuLys------LysLysAlaLysSerGluSerGluSerGluSerGluAspGluL 216

Qy     852 GAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAA 911
             ::|||   |||         :::|||   ||||||          |||      :::
Db     216 ysGluValLysLysSerLysLysLysSerLysLysValValLysLysGluSerGluSerG 236

Qy     912 GTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTG 971
               :::      |||                |||         |||         ||||
Db     236 luAspGluAlaProGluLysLysLysThrGluLysArgLysArgSerLysThrSerSerG 256

Qy     972 AGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAG 1031
             ||:::                  |||:::|||   :::            :::|||
Db     256 luGluSerSerGluSerGluLysSerAspGluGlu---------GluGlu---------- 269

Qy    1032 GAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC-----------GAATCACCGATGG 1079
                           ||||||||||||         |||             :::   |||:::
Db     270 ----------------GluLysGluSerSerProLysProLysLysLysLysProLeuA 284

Qy    1080 ATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCA- 1138
              |||            :::         :::      |||:::|||   |||   |||
Db     284 laValLysLysLeuSerSerAspGluGluSerGluGluSerAspValGluValLeuProG 304

Qy    1139 ------------------------------------GGTTTTGATGAACAAGAAGATG 1160
                                                      ||||||::::::
Db     304 lnLysLysLysArgGlyAlaValThrLeuIleSerAspSerGluAspGluLysAspGlnL 324

Qy    1161 GTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTG 1220
               |||   |||:::   ::::::          :::         :::   |||
Db     324 ysSerGluSerGluAlaSerAspValGluGluLysValSerLysLysLysAla------- 341

Qy    1221 AATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAG 1280
                 :::|||:::   |||::::::                                   |
Db     342 -----LysLysGlnGluSerSerGlu---------------------------SerG 350

Qy    1281 GAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCTA 1340
             ||      |||      :::|||   |||                                |
Db     350 lySerAspSerSerGluGlySerIleThrValAsnArgLys------------------S 364

Qy    1341 GCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA------AAAAAGGATTCTAAAA 1394
             ||         |||         |||:::   |||                ||||||:
Db     364 erLysLysLysGluLysProGluLysLysLysLysGlyIleIleMetAspSerSerLysL 384

Qy    1395 TCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAA 1454
             ::         :::            :::||||||      |||:::::::::   |||
Db     384 euGlnLysGluThrIleAspAlaGluArgAlaGluLysGluArgArgLysArgLeuGluL 404

Qy    1455 CCAAACATCAAAGA--------------------------------------------- 1468
             |||   :::
Db     404 ysLysGlnLysGluPheAsnGlyIleValLeuGluGluGlyGluAspLeuThrGluMetL 424

Qy    1469 --------------ACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCTCCAG 1514
                  ::::::|||||:::            |||         |||   ||||
Db     424 euThrGlyThrSerSerGlnArgLysLeuLysSerValVal---------LeuAspProA 441

Qy    1515 ATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCC 1574
```

```
              ||                :::  ::::::|||      :::  |||
Db      441 sp------------SerSerThrValAspGluGluSerLysLysPro------------- 452

Qy     1575 GGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGC 1634
                                                  |||   :::      |||:::|
Db      453 ----------------------------------ValGluValHisAsnSerLeuValA 461

Qy     1635 GAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGATAACAAGC 1694
             ||          :::  ||||||         |||
Db      461 rg---------IleLeuLysProHisGlnAlaHis------------------------ 469

Qy     1695 ATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGC 1754
                                   |||   :::|||            |||:::   :::|
Db      470 -----------------GlyIleGlnPheMetTyrAspCysAlaCysGluSerLeuAspA 484

Qy     1755 GA------ACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTA 1808
             ||     |||   ||||||
Db      484 rgLeuAspThrGluGlySer--------------------------------------- 490

Qy     1809 CACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAAT 1868
                                       ||||||   :::   :::           |||
Db      491 --------------------GlyGlyIleLeuAlaHisCysMetGlyLeuGlyLysThr- 503

Qy     1869 GGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAAC 1928
                           :::|||   |||        ||||||
Db      504 --------LeuGlnValIleThrPheLeuHisThrValLeu------------------ 514

Qy     1929 ATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCA---AATA 1985
              :::::::::    |||                  ::::::   :::   :::|||   |||:
Db      515 --MetHisGluLysIleGlyGluLysCysLysArgValLeuValValValProLysAsnV 534

Qy     1986 TTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGC 2045
            ::                          ::::::               :::      :::
Db      534 alIleIleAsnTrpPhe------------LysGluPheGlnLysTrpLeuValAspAsnA 550

Qy     2046 AGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATG 2105
              |||::::::   ::::::      |||
Db      550 spGluGluLeuAspThrIleAspValAsn------------------------------ 559

Qy     2106 CTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGA------TTGT 2159
                           |||      ||||||      |||         |||     |||
Db      560 --------------GluLeuAspSerTyrLysThrIleGluAspArgArgArgAlaLeuL 575

Qy     2160 TTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAA 2219
                     ||||||:::   |||      :::           :::   ||||||
Db      575 ysAlaTrpHisSerSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgI 595

Qy     2220 GAGTCACA------------------GAGAAAAAACCTACTGGGTTGGTGACATTTACAA 2261
              :::|||                   :::|||||||||
Db      595 leLeuThrValGluAspAspProLysLysLysLysProLysAsnArgAsnArgArgLeuG 615

Qy     2262 GACAGAGTCTTGAAGATTTT-----------------------CCAGAATGGGAAAGAT 2297
              ::::::   |||||||||                        |||:::        |
Db      615 luLysAlaLysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValC 635

Qy     2298 GTGAA---------------------------------------------------AAAC 2306
             ||:::                                                   |||
Db      635 ysAspGluAlaHisLysLeuLysAsnAspAspSerAlaLeuSerLysCysMetValLysI 655

Qy     2307 CCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGC---------- 2356
              ||||||:::      :::            ||||||      :::|||
Db      655 leLeuThrLysArgArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluT 675

Qy     2357 --------------------------CAAGGCATGCTA------CAGGTGGATTTTGCAA 2384
                                       |||:::|||      :::   :::|||||||
Db      675 yrHisCysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaA 695

Qy     2385 ATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAA------------G 2432
              ||||||||||        :::   :::   |||      ::::::               |
Db      695 snArgPheVal------AsnIleIleAsnArgGlyArgThrLysAspAlaSerProLeuG 713

Qy     2433 AAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGG 2492
             ||:::   |||:::                 ::::::              :::  :::|
Db      713 luValSerPheMetLysArgArgCysHisValLeuTyrAspHisLeuLysLysCysValA 733

Qy     2493 ATCACAATGAATGTCTAATTATCACAGGTACT---------------------------- 2524
             ||        :::        ::::::|||
Db      733 spArgLysAspTyrArgValLeuThrGluAlaIleProProLysGlnGluTyrValIleA 753

Qy     2525 -----------------GAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTT 2567
                             |||   :::   |||   |||   |||         ||||||
Db      753 snValArgGlnThrGluArgGlnCysAlaLeuTyr-AsnAlaPheLeuAsnAspIleVal 772

Qy     2568 GGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGA 2627
             |||                                           ||||||   :::   |||
Db      773 GlyAsp-------------------------------------SerGlyLeuSerLysArg 780

Qy     2628 TCGTTGCCA 2636
               ||||||
Db      781 LeuLeuPro 783

RESULT 3
PCLO_RAT
ID   PCLO_RAT       STANDARD;      PRT;  5085 AA.
AC   Q9JKS6; Q9JLT1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN   PCLO.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX   MEDLINE=20170257; PubMed=10707984;
RA   Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA   Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT   "Piccolo, a presynaptic zinc finger protein structurally related to
RT   bassoon.";
RL   Neuron 25:203-214(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP   VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP   ALA-4694.
RX   MEDLINE=21181819; PubMed=11285225;
RA   Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT   "An unusual C(2)-domain in the active-zone protein piccolo:
RT   implications for Ca(2+) regulation of neurotransmitter release.";
RL   EMBO J. 20:1605-1619(2001).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC       junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9JKS6-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF138789; AAF07822.2; -.
DR   EMBL; AF227534; AAF63196.1; -.
DR   HSSP; P04410; 1A25.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; IDA.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN       372    491       12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                 P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING      523    547       C4-TYPE (POTENTIAL).
FT   ZN_FING     1010   1033       C4-TYPE (POTENTIAL).
FT   DOMAIN      2351   2362       POLY-PRO.
FT   DOMAIN      4442   4536       PDZ.
FT   DOMAIN      4653   4752       C2 DOMAIN 1.
FT   DOMAIN      4968   5059       C2 DOMAIN 2.
FT   VARSPLIC    4876   4880       TKPTN -> SKRRK (in isoform 2).
FT                                 /FTId=VSP_003930.
FT   VARSPLIC    4881   5085       Missing (in isoform 2).
FT                                 /FTId=VSP_003931.
FT   MUTAGEN     4668   4668       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                 AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                 BINDING ACTIVITY.
FT   MUTAGEN     4674   4674       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                 AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                 BINDING ACTIVITY.
FT   MUTAGEN     4688   4688       V->S: SMALL INCREASE IN AFFINITY FOR
FT                                 CALCIUM.
FT   MUTAGEN     4688   4689       VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                 CALCIUM.
FT   MUTAGEN     4689   4689       M->S: INCREASED AFFINITY FOR CALCIUM.
FT   MUTAGEN     4690   4691       VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                 CALCIUM.
FT   MUTAGEN     4692   4693       QN->AA: MODERATE INCREASE IN AFFINITY FOR
FT                                 CALCIUM.
FT   MUTAGEN     4694   4694       A->S: NO EFFECT ON CALCIUM-BINDING
FT                                 ACTIVITY.
SQ   SEQUENCE   5085 AA;  552702 MW;  5A1BB543201A7450 CRC64;

Alignment Scores:
Pred. No.:                0.000296       Length:        5085
Score:                    181.50         Matches:       150
Percent Similarity:       36.39%         Conservative:  128
Best Local Similarity:    19.63%         Mismatches:    269
Query Match:              2.52%          Indels:        217
DB:                       1              Gaps:          35

US-09-302-812-3 (1-4069) x PCLO_RAT (1-5085)

Qy         63 GCCAAGGCAGGCGCTGAGAAGGACGCGCAGTCCATCTCTCTCAGGTTAGTGAAATGAGGC 122
                |||||||||||   |||:::
Db       1232 AlaLysAlaGlyGluGluGluGlnGlnProAsp-----------------------Ser 1243

Qy        123 TCTCCGCGGGGCCGGCCCGGGGACAGTGCGCTGCTGGTCCCAGCATGAATGCGGGCCCCG 182
                 |||         ||||||   :::    |||    :::|||            :::|||
Db       1244 ArgProGluAlaLeuProGlyAlaThrProLeuThrLeuPro---------LysAlaGly 1260

Qy        183 GCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTT 242
                             |||    :::|||
Db       1261 GluLysGluArgAlaValAlaGlnPro--------------------------------- 1269

Qy        243 CGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACG------ 296
               :::      |||:::   :::   |||:::|||      |||        :::|||
Db       1270 GlnAlaGluGlySerSerLysAspGlyGlnGlyGluArgSerLysGluLysThrGluLys 1289

Qy        297 ---------CTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCT------------ 335
                        :::   :::|||   |||   :::|||:::   |||
Db       1290 GluGluAspLysSerAspThrSerSerSerGlnGlnProLysSerProGlnGlyLeuSer 1309

Qy        336 ------------GCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTG 383
                          :::|||   |||            |||            |||      :::
Db       1310 AspThrGlyTyrSerSerAspGlyIleSerGlySerLeuGlyGluIleProSerLeuIle 1329

Qy        384 TTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAAT 443
                 |||:::::::::      |||      |||      ||||||:::      :::
Db       1330 ProSerAspGluLysAspLeuLeuLysGly------LeuLysLysAsp--SerPheSerG 1347

Qy        444 CAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTG 503
                |||||      |||         ::::::         ::::::||||||   :::   |||:
Db       1347 lnGluSerSerProSerSerProSerAspLeuAlaLysLeuGluSerThrValLeuSerI 1367

Qy        504 TA---------CAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTT 554
               ::          |||                              ||||||
Db       1367 leLeuGluAlaGlnAlaSerThrLeuValGlyGluLysAlaGluLysLysThrGlnProG 1387

Qy        555 CTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATC 614
               ::::::|||     :::|||            ::::::   ||||||     ::::::
Db       1387 lnLysIleSerProGluLysProGlnAspGlnGlnLysThrGlnThrAlaSerGluThrL 1407

Qy        615 AGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAA 674
                         :::   :::                          |||::::::   ||||
Db       1407 euAspIleThrIleSerGlu------------------------GluGluIleLysGluS 1419

Qy        675 GTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATT------- 727
               |||:::   :::   |||:::          |||::::::   |||
Db       1419 erGlnGluLysLysValSer------ProLysLysAspSerGluGlnGlyPheProSerA 1437

Qy        728 -----------------------------------GATCGGTCACCTCAAAATGATGATC 752
                                                  |||   ||||||:::         :
Db       1437 rgLysGluHisLysGluLysProGluLeuValAspAspLeuSerProArgArgAlaSerT 1457

Qy        753 ACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGC 812
               ::        :::                  :::|||         :::         :::
Db       1457 yrAspSerValGluAspSerSerGluSerGluAsnSerProValValArgArgLysArgA 1477

Qy        813 TTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGA 872
                   :::         ::::::   ||||||:::   :::|||
Db       1477 rgThrSerIleGlySerSerSerSerAspGluTyrLysGlnGlu---------------- 1491

Qy        873 AGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAG 932
                          |||      ||||||                       |||     |
Db       1492 --------AspSerGlnGlySerGlyGlu-----------------------GluGluA 1501

Qy        933 ACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAA 992
               ||   :::   |||      :::   ::::::   :::::::::         :::||||||:
Db       1501 spPheIleArgLysGln---IleIleGluMetSerAlaAspGluAspAlaSerGlySerG 1520

Qy        993 AAAATGACAACAAATTGATTAGACAAGAA------------AGTTGCCTAGGAAATTCTC 1040
               ::   |||   :::   ||||||   :::                  |||   :::|||   |||
Db       1520 lu---AspGluGluPheIleArgSerGlnLeuLysGluIleSerGlyValGlyGluSerG 1539

Qy       1041 CTCCATTTGAGAAGGAAAGT---------------------------------------- 1060
                       |||:::
Db       1539 lnLysArgGluGluAlaLysGlyLysGlyLysGlyValAlaGlyLysHisArgArgLeuT 1559

Qy       1061 -----GAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAG 1115
```

```
                  :::      :::       |||   |||       :::                |
Db     1559 hrArgLysSerSerThrSerPheAspAspAspAlaGlyArgArgHisSerTrpHisAspG 1579

Qy     1116 AAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAA 1175
             ||    |||:::||||||              ||||||||     :::          :::|
Db     1579 lu---AspAspGluThr---------PheAspGluSerProGluLeuLysPheArgGluT 1595

Qy     1176 CAGCAAATAAACCTTCA-----------------------------------AGGTTCC 1199
             ||    ::::::   |||                                  ||||||:
Db     1595 hrLysSerGlnGluSerGluGluLeuValValAlaGlyGlyGlyGlyLeuArgArgPheL 1615

Qy     1200 AAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTA 1259
             ::      :::    :::                  :::||||||::::::      :::   :
Db     1615 ysThrIleGluLeuAsnSerThrIleAlaAspLysTyrSerSerGluSerSerGlnLysL 1635

Qy     1260 GA------TTACATTTCCAATTTGAA-----------------------------GGAG 1283
             ::         |||:::|||      |||
Db     1635 ysThrIleLeuTyrPheAspGluGluProGluLeuGluMetGluSerLeuThrAspSerP 1655

Qy     1284 GAGAGAGTCGCACT---------GGAATGAATGATTTAAATGCTAAA------------- 1321
               |||    |||:::          |||    :::   |||:::|||
Db     1655 roGluAspArgSerArgGlyGluGlySerSerSerLeuHisAlaSerSerPheThrProG 1675

Qy     1322 -----CTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATG 1376
                       |||    ::::::|||||||||:::   :::      :::|||      |||
Db     1675 lyThrSerProThrSerValSerSerLeuAspGluAspSerAspSerSerProSerHisL 1695

Qy     1377 GAAAAAAGGATTCTAAA------------------ATCACAGATCATTTGATGAGACTGC 1418
               |||    :::||||||                          |||        ||||
Db     1695 ysLysGlyGluSerLysGlnGlnArgLysAlaArgHisArgSerHisGlyProLeuLeuP 1715

Qy     1419 CCAAAGCAGAGGACAGA------------------AGAAAAGAACAGTGGGAAACCAAAC 1460
             ||        ||||||                      |||:::|||:::          |||
Db     1715 roThrIleGluAspSerSerGluGluGluGluLeuArgGluGluGluLeuLeuLysG 1735

Qy     1461 ATCAAAGAACAGAAAGGAAGATC------------------------------------- 1483
               |||        :::|||::::::
Db     1735 luGlnGluLysGlnArgGluLeuGluGlnGlnGlnArgLysSerSerSerLysLysSerL 1755

Qy     1484 ----------------------------------------------CCTAAATACGTTCCAC 1499
                                                         ||||||      |||
Db     1755 ysLysAspLysAspGluLeuArgAlaGlnArgArgArgGluArgProLysThrProProS 1775

Qy     1500 CTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGC 1559
               :::|||||||||     :::                 :::|||    ||||||:::|||:::
Db     1775 erAsnLeuSerProIleGluAsp------AlaSerProThrGluGluLeuArgGlnAlaA 1793

Qy     1560 CTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT---GCCAATCACACAGTAACTA 1616
                                       |||    |||    |||     :::   ::::::   :::|
Db     1793 laGluMetGlu---------GluLeuHisArgSerSerCysSerGluTyrSerProSerI 1810

Qy     1617 TTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAG 1676
             ||        |||                            |||:::    |||            :::
Db     1810 leGluSerAsp--------------------ProGluGlyPheGluIleSerProGluL 1823

Qy     1677 ATTTGTGGGAT---AACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAG 1733
               :::    :::      |||         |||:::|||    :::                  |||
Db     1823 ysIleIleGluValGlnLysValTyrLysLeuProAlaAlaValSerLeuTyrSerProT 1843

Qy     1734 TGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCAC 1793
               :::::::::::      :::          |||                  :::|||    |||
Db     1843 hrAspGluGlnSerValMetGlnLysGluGly---------------ValGlnLysAla- 1857

Qy     1794 TTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGG 1853
                                           ||||||    |||         |||
Db     1858 ----------------------------LeuLysSerAlaGluGluMetTyrGluGluM 1868

Qy     1854 CATATTCTAAGAAATGGGACTTTACAGCTTTG------ATCGATTTCTGGGATAAGGTAC 1907
                    |||          :::   |||              :::        |||:::|||
Db     1868 etMetGlnLysProHisLysTyrLysAlaPheProAlaAlaAsnGluArgAspGluValP 1888

Qy     1908 TTGAAGAAGCAGAA---------------------GCTCAACATTTATATCAGTCCATCT 1946
               |||:::                                  ::::::|||:::|||::::
Db     1888 heGluLysGluProLeuTyrGlyGlyMetLeuIleGluAspTyrIleTyrGluSerLeuV 1908

Qy     1947 TGCCTGAT 1954
               ::    |||
Db     1908 alGluAsp 1910
```

```
RESULT 4
PCLO_HUMAN
ID   PCLO_HUMAN     STANDARD;      PRT;  5147 AA.
AC   Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragments).
GN   PCLO OR ACZ OR KIAA0559.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 4405-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
```

```
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9Y6V0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC                                  VSP_003926, VSP_003927;
CC         Note=No experimental confirmation available;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19188; CAB60727.1; -.
DR   EMBL; AC004903; AAD20936.1; -.
DR   EMBL; AC004886; AAD21789.1; -.
DR   EMBL; AB011131; BAA25485.1; -.
DR   EMBL; BC001304; AAH01304.1; -.
DR   EMBL; AC004082; AAB97937.1; -.
DR   PIR; T00634; T00634.
DR   HSSP; P04410; 1A25.
DR   Genew; HGNC:13406; PCLO.
DR   MIM; 604918; -.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0045202; C:synaptic junction; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001565; Synaptotagmin.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
KW   Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN      400    465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-P-X.
FT   ZN_FING     499    523       C4-TYPE (POTENTIAL).
FT   ZN_FING     969    992       C4-TYPE (POTENTIAL).
FT   NON_CONS   1010   1011
FT   DOMAIN     2300   2325       POLY-PRO.
FT   DOMAIN     4391   4442       PDZ.
FT   DOMAIN     4544   4633       C2 DOMAIN 1.
FT   DOMAIN     5031   5121       C2 DOMAIN 2.
FT   VARSPLIC   4404   4404       S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                                QTGKLMEG (in isoform 2).
FT                                /FTId=VSP_003923.
FT   VARSPLIC   4534   4534       K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
FT                                /FTId=VSP_003924.
FT   VARSPLIC   4576   4576       G -> GQVMVVQNAS (in isoform 2).
FT                                /FTId=VSP_003925.
FT   VARSPLIC   4757   4761       TAHKS -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003926.
FT   VARSPLIC   4762   5147       Missing (in isoform 2).
FT                                /FTId=VSP_003927.
SQ   SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

Alignment Scores:
Pred. No.:              0.000945        Length:         5147
Score:                  174.00          Matches:        140
Percent Similarity:     36.67%          Conservative:   120
Best Local Similarity:  19.75%          Mismatches:     251
Query Match:            2.41%           Indels:         198
DB:                     1               Gaps:           32

US-09-302-812-3 (1-4069) x PCLO_HUMAN (1-5147)

Qy        191 CCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACAACTTCGCCGGCTGCTTCGGACGCC 250
              |||||||||                        |||                   |||
Db       1227 ProCysThr---------------------AlaLysProAspGlnGluLysGluAspAsp 1239

Qy        251 CGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCACGTGCAGTTC 310
              :::|||      |||        ::::::
Db       1240 LysSerAspThrSerSerSerGlnGln--------------------------------- 1248

Qy        311 AGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGC 370
                        |||  ||||||        :::                        |||
Db       1249 ---------ProLysSerProGlnGlyLeuSerAspThrGlyTyrSerSerAspGlyIle 1265

Qy        371 GCCACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT--------- 421
              ::::::||||||                |||  |||  :::  |||
Db       1266 SerSerSerLeu------------GlyGluIleProSerLeuIleProThrAspGluLys 1281

Qy        422 ---------AAAGGAATCAAG------ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
                       ||||||:::|||      :::         ||||||      |||
Db       1282 AspIleLeuLysGlyLeuLysLysAspSerPheSerGlnGluSerSerProSerSerPro 1301

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ::::::      ::::::||||||   :::   |||:::    :::
Db       1302 SerAspLeuAlaLysLeuGluSerThrValLeuSerIleLeuGlu--------------- 1316

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
                                 |||  |||:::    :::||||||        |||
Db       1317 ------------------AlaGlnAlaSerThrLeuAlaAspGluLysSer------Glu 1328

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              |||   ||||||           ||||:::  :::      :::          :::
Db       1329 LysLysThrGln---------ProHisGluValSerPro------GluGlnProLysAsp 1343

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
                 :::  |||:::  |||      ||||||         :::||||:::      |||
Db       1344 GlnGluLysThrGlnSerLeu------SerGluThrLeuGluIleThrIleSerGluGlu 1361

Qy        707 CAGTTTAGTAATGCTAACATTGATCGG------------------TCACCTCAAAATGAT 748
              :::      :::      :::|||                        |||   |||
Db       1362 GluIleLysGluSerGlnGluGluArgLysAspThrPheLysLysAspSerGlnGlnAsp 1381

Qy        749 GATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTA 808
                 |||         |||   :::|||       :::             :::||||||
Db       1382 IleProSerSerLysAspHisLysGluLysSerGluPheValAspAspIleThrThrArg 1401

Qy        809 AAG------CTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCC------------ 850
              :::            :::   ::::::::::::|||   |||::::::
Db       1402 ArgGluProTyrAspSerValGluGluSerSerGluSerGluAsnSerProValProGln 1421

Qy        851 AGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCA 910
              |||:::   ::::::                 |||   |||          ::::::   :::
Db       1422 ArgLysArgArgThrSerValGlySerSerSerSerAspGluTyrLysGlnGluAspSer 1441

Qy        911 AGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCT 970
                       :::|||   |||   :::   |||         :::   ::::::   :::
Db       1442 GlnGlySerGlyGluGluGluAspPheIleArgLysGln---IleIleGluMetSerAla 1460

Qy        971 GAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTA 1030
              ::::::         :::||||||::::::|||        :::         |||      :::
Db       1461 AspGluAspAlaSerGlySerGluAspAspGluPheIleArgAsnGlnLeuLysGluIle 1480
```

```
Qy       1031 GGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAA--------------------- 1069
                 :::|||       :::|||:::   |||   :::
Db       1481 SerSerSerThrGluSerGlnLysLysGluGluThrLysGlyLysGlyLysIleThrAla 1500

Qy       1070 ---------------------------------TCACCGATGGATGTGGATAATTCTAAA 1096
                                               :::  :::|||  |||      :::
Db       1501 GlyLysHisArgArgLeuThrArgLysSerSerThrSerIleAspGluAspAlaGlyArg 1520

Qy       1097 AATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAA 1156
                                  |||  ||||||           ||||||||
Db       1521 ArgHisSerTrpHisAspGluAspAspGlu------------AlaPheAspGluSerPro 1536

Qy       1157 GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCA------------------------ 1192
               :::              :::|||  ::::::  |||
Db       1537 GluLeuLysTyrArgGluThrLysSerGlnGluSerGluGluLeuValValThrGlyGly 1556

Qy       1193 ------------AGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCT 1240
                          ||||||:::      :::   :::              :::||||||
Db       1557 GlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrIleAlaAspLysTyrSer 1576

Qy       1241 ACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTT------------------------ 1276
                 :::     :::   :::             |||
Db       1577 AlaGluSerSerGlnLysLysThrSerLeuTyrPheAspGluGluProGluLeuGluMet 1596

Qy       1277 ------------GAAGGAGGAGAGAGTCGCACT---------GGAATGAATGATTTAAAT 1315
                          :::      |||   |||:::       |||   :::   |||:::
Db       1597 GluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlySerSerSerLeuHis 1616

Qy       1316 GCTAAA------------------CTACCTGGAAATATTTCTAGCCTGAATGTAGAATGC 1357
              |||                     |||   ::::::|||||||||:::   :::
Db       1617 AlaSerSerPheThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSer 1636

Qy       1358 AGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAA------------------ATCACA 1399
                 :::|||     |||   |||   :::||||||
Db       1637 AspSerSerProSerHisLysLysGlyGluSerLysGlnGlnArgLysAlaArgHisArg 1656

Qy       1400 GATCATTTGATGAGACTGCCCAAAGCAGAGGACAGA------------------AGAAAA 1441
                 |||         ||||||      ||||||                 |||:::
Db       1657 ProHisGlyProLeuLeuProThrIleGluAspSerSerGluGluGluGluLeuArgGlu 1676

Qy       1442 GAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC------------------ 1483
              |||:::         |||   |||      :::|||:::|||
Db       1677 GluGluGluLeuLeuLysGluGlnGluLysGlnArgGluIleGluGlnGlnGlnArgLys 1696

Qy       1483 ------------------------------------------------------------ 1483

Db       1697 SerSerSerLysLysSerLysLysAspLysAspGluLeuArgAlaGlnArgArgArgGlu 1716

Qy       1484 ---CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
                 ||||||      |||   :::|||||||||   :::            :::|||
Db       1717 ArgProLysThrProProSerAsnLeuSerProIleGluAsp------AlaSerProThr 1734

Qy       1541 GAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT--- 1597
              ||||||:::|||:::                           |||   |||   |||
Db       1735 GluGluLeuArgGlnAlaAlaGluMetGlu---------GluLeuHisArgSerSerCys 1751

Qy       1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657
              :::   ::::::   :::|||      |||                       |||:::
Db       1752 SerGluTyrSerProSerIleGluSerAsp---------------------ProGluGly 1764

Qy       1658 TTTCCAACACATTATAAAGATTTGTGGGAT---AACAAGCATGTTAAAATGCCTTGTTCA 1714
              |||         :::   :::   :::         |||      |||:::|||   :::
Db       1765 PheGluIleSerProGluLysIleIleGluValGlnLysValTyrLysLeuPro---Thr 1783

Qy       1715 GAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGG 1774
                    :::||||||         ||||||
Db       1784 AlaValSerLeuTyrSerProThrAspGluGln--------------------------- 1794

Qy       1775 GAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
                             :::::::::   |||            :::   ||||||   |||
Db       1795 ---------------SerIleMetGlnLysGluGlySerGlnLysAlaLeuLysSerAla 1809

Qy       1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTG------ATC 1888
                          |||          |||            :::   |||
Db       1810 GluGluMetTyrGluGluMetMetHisLysThrHisLysTyrLysAlaPheProAlaAla 1829

Qy       1889 GATTTCTGGGATAAGGTACTTGAAGAAGCAGAA---------------------GCTCAA 1927
              :::         |||:::|||   |||:::
Db       1830 AsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAsp 1849

Qy       1928 CATTTATATCAGTCCATCTTGCCTGAT 1954
              ::::::|||:::|||::::::   |||
Db       1850 TyrIleTyrGluSerLeuValGluAsp 1858
```

RESULT 5
PCLO_MOUSE

```
ID   PCLO_MOUSE     STANDARD;      PRT;  5038 AA.
AC   Q9QYX7; Q9QYX6; Q9QZJ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-
DE   derived HLMN protein).
GN   PCLO OR ACZ.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RP   TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   REVISIONS.
RA   Kilimann M.W.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 4502-4682 FROM N.A.
RC   TISSUE=Brain;
RA   Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   INTERACTION WITH RIMS2.
RX   MEDLINE=22384373; PubMed=12401793;
RA   Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA   Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT   "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT   cAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis.";
RL   J. Biol. Chem. 277:50497-50502(2002).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking.
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin.
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9QYX7-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC       stomach. Not detected in other tissues analyzed including adrenal
CC       gland, testis and pancreas.
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
```

```
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19185; CAB60731.2; -.
DR   EMBL; Y19186; CAB60732.2; -.
DR   EMBL; AF181269; AAD55786.2; -.
DR   HSSP; P04410; 1A25.
DR   MGD; MGI:1349390; Pclo.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; IDA.
DR   GO; GO:0019933; P:cAMP-mediated signaling; IDA.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
DR   GO; GO:0030073; P:insulin secretion; IDA.
DR   GO; GO:0017157; P:regulation of exocytosis; IDA.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      371    470       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     502    526       C4-TYPE (POTENTIAL).
FT   ZN_FING     967    990       C4-TYPE (POTENTIAL).
FT   DOMAIN     2305   2329       POLY-PRO.
FT   DOMAIN     4394   4488       PDZ.
FT   DOMAIN     4607   4705       C2 DOMAIN 1.
FT   DOMAIN     4922   5012       C2 DOMAIN 2.
FT   VARSPLIC   4829   4833       TKPTN -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003928.
FT   VARSPLIC   4834   5038       Missing (in isoform 2).
FT                                /FTId=VSP_003929.
SQ   SEQUENCE   5038 AA;  547600 MW;  DADA460CF3B40888 CRC64;

Alignment Scores:
Pred. No.:              0.00119         Length:         5038
Score:                  172.50          Matches:        136
Percent Similarity:     37.55%          Conservative:   125
Best Local Similarity:  19.57%          Mismatches:     251
Query Match:            2.39%           Indels:         183
DB:                     1               Gaps:           30

US-09-302-812-3 (1-4069) x PCLO_MOUSE (1-5038)

Qy        252 GGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACG--------------- 296
               |||      :::   ||||||||        |||       ||||||
Db       1231 GlyThrAlaLysGluGlyArgGlyGluProSerLysAspArgThrGluLysGluGluAsp 1250

Qy        297 CTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCT--------------------- 335
                 :::    :::|||    |||    :::|||:::    |||
Db       1251 LysSerAspThrSerSerSerGlnGlnProLysSerProGlnGlyLeuSerAspThrGly 1270

Qy        336 ---GCGTCCCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAAC 392
                 :::|||   |||          |||          |||     :::   |||:::
Db       1271 TyrSerSerAspGlyIleSerGlySerLeuGlyGluIleProSerLeuIleProSerAsp 1290

Qy        393 AAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTT 452
              ::::::   |||         |||      ||||||:::        :::       ||||||
Db       1291 GluLysAspLeuLeuLysGly------LeuLysLysAsp--SerPheSerGlnGluSerS 1308

Qy        453 TGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTA------- 505
                 |||      ::::::      ::::::||||||   :::    |||:::
Db       1308 erProSerSerProSerAspLeuAlaLysLeuGluSerThrValLeuSerIleLeuGluA 1328

Qy        506 --CAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAA 563
                 |||                                |||||               ::::::|
Db       1328 laGlnAlaSerThrLeuValGlyGluLysAlaGluLysLysThrGlnProGlnLysValS 1348

Qy        564 GTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAG 623
               ||    ::::::         ::::::   ||||||       ::::::   :::        :
Db       1348 erProGluGlnProGlnAspGlnGlnLysThrGlnThrProSerGluThrArgAspIleS 1368

Qy        624 CAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTC 683
               :::::    :::                                  |||::::::   ||||||:::    :
Db       1368 erIleSerGlu------------------------GluGluIleLysGluSerGlnGluL 1380

Qy        684 AAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAA 743
               ::      ||||||                                      |||     ::::
Db       1380 ysLysValThr---------------------------------------SerLysLysA 1387

Qy        744 ATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAA 803
               ::                        :::|||:::::::::|||    :::    :::
Db       1387 spSerAlaGlnGlyPheProSerArgLysGluHisLysGluAsnProGluLeuValAspA 1407

Qy        804 CTGTAAAGCTTGCAAATGCA---------------AAGCAGACTACGGAAGATGAACACG 848
                 :::            |||            :::     ::::::|||     |||::::
Db       1407 spLeuSerProArgArgAlaSerTyrAspSerValGluAspSerSerGluSerGluAsnS 1427

Qy        849 CCAGAGAAGCCAAAAGCCACCAGAAG---------------TGCAGCAAGTCTTGCCATC 893
               ::          |||:::        ::::::               |||          :::
Db       1427 erProValAlaArgArgLysArgArgThrSerIleGlySerSerSerGluGluTyrL 1447

Qy        894 CTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCAT 953
                 ||||||                     :::||||||     |||     :::     |||        :
Db       1447 ysGlnGluAspSerGlnGlySerGlyGluAspGlu---AspPheIleArgLysGln---I 1465

Qy        954 TGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTA 1013
               ::     ::::::     :::::::::          :::||||||:::::::::          :::
Db       1465 leIleGluMetSerAlaAspGluAspAlaSerGlySerGluAspGluGluPheIleArgS 1485

Qy       1014 GACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC------- 1066
                 |||          :::|||                 :::|||          |||
Db       1485 erGlnLeuLysGluIleGlyGlyValThrGluSerGlnLysArgGluGluThrLysGlyL 1505

Qy       1067 -----GAATCACCGATGGAT---------------GTGGATAATTCTAAAAATAGTTGTC 1106
                    :::||||||                         |||          |||
Db       1505 ysGlyLysSerProAlaGlyLysHisArgArgLeuThrArgLysSerSerThrSerPheA 1525

Qy       1107 AAGACTCA-----------------------GAAGCAGATGAGGAGACAAGTCCAGGTT 1142
               |||                            :::     |||:::||||||               |
Db       1525 spAspAspAlaGlyArgArgHisSerTrpHisAspGluAspAspGluThr---------P 1542

Qy       1143 TTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCA---------- 1192
              |||||||||          :::               :::|||     ::::::     |||
Db       1542 heAspGluSerProGluLeuLysPheArgGluThrLysSerGlnGluSerGluGluLeuV 1562

Qy       1193 --------------------------AGGTTCCAAGCAAGAGACGCTGACATTGAATTTA 1226
                                          ||||||:::          :::     :::
Db       1562 alValAlaGlyGlyGlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrValT 1582

Qy       1227 GGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTT---------- 1276
                 :::||||||     :::          :::     :::               |||
Db       1582 hrAspLysTyrSerAlaGluSerSerGlnLysLysThrThrLeuTyrPheAspGluGluP 1602

Qy       1277 -------------------------GAAGGAGGAGAGAGTCGCACT---------GGAA 1301
                                          :::          |||     |||:::               |||
Db       1602 roGluLeuGluMetGluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlyS 1622

Qy       1302 TGAATGATTTAAATGCTAAA------------------CTACCTGGAAATATTTCTAGCC 1343
```

```
                :::   |||:::|||                       |||   ::::::||||||
Db       1622 erSerSerLeuHisAlaSerSerPheThrProGlyThrSerProThrSerValSerSerL 1642

Qy       1344 TGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAA---------- 1393
              ||:::   :::      :::|||      |||   |||   :::||||||
Db       1642 euAspGluAspSerAspSerSerProSerHisLysLysGlyGluSerLysGlnGlnArgL 1662

Qy       1394 --------ATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGA---------- 1435
                          |||          ||||||     ||||||
Db       1662 ysAlaArgHisArgSerHisGlyProLeuLeuProThrIleGluAspSerSerGluGluG 1682

Qy       1436 --------AGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC---- 1483
                      |||:::|||:::           |||   |||     :::|||::::::
Db       1682 luGluLeuArgGluGluGluGluLeuLeuLysGluGlnGluLysGlnArgGluLeuGluG 1702

Qy       1483 ------------------------------------------------------------ 1483

Db       1702 lnGlnGlnArgLysSerSerSerLysLysSerLysLysAspLysAspGluLeuArgAlaG 1722

Qy       1484 -----------------CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGC 1526
                               ||||||     |||   :::|||||||||   :::
Db       1722 lnArgArgArgGluArgProLysThrProProSerAsnLeuSerProIleGluAsp---- 1740

Qy       1527 TTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCT 1586
                 :::|||   ||||||:::|||:::                          |||
Db       1741 --AlaSerProThrGluGluLeuArgGlnAlaAlaGluMetGlu---------GluLeuH 1757

Qy       1587 TGAGACCATCT---GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAG 1643
                |||   |||   :::   ::::::   :::|||      |||
Db       1757 isArgSerSerCysSerGluTyrSerProSerIleGluSerAsp---------------- 1771

Qy       1644 AAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGAT---AACAAGCATGTTA 1700
                     |||:::   |||          :::   :::   :::      |||      |
Db       1772 -----ProGluGlyPheGluIleSerProGluLysIleIleGluValGlnLysValTyrL 1790

Qy       1701 AAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTG 1760
              ||:::|||   :::      :::||||||         ||||||
Db       1790 ysLeuPro---ThrAlaValSerLeuTyrSerProThrAspGluGln------------- 1804

Qy       1761 CGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAA 1820
                                           :::::::::   |||         :::
Db       1805 -----------------------------SerValMetGlnLysGluGlyAlaGlnLysA 1815

Qy       1821 ACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAG 1880
                ||||||   |||         |||               |||        :::   |
Db       1815 laLeuLysSerAlaGluGluMetTyrGluGluMetMetHisLysProHisLysTyrLysA 1835

Qy       1881 CTTTG------ATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAA------------- 1921
              ||                :::      |||:::|||   |||:::
Db       1835 laPheProAlaAlaAsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyM 1855

Qy       1922 --------GCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                              ::::::|||:::|||::::::   |||
Db       1855 etLeuIleGluAspTyrIleTyrGluSerLeuValGluAsp 1868
```

```
RESULT 6
ATRX_PANTR
ID   ATRX_PANTR     STANDARD;      PRT;  2492 AA.
AC   Q7YQM4;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP).
GN   ATRX.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB102642; BAC81111.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING     220    268       PHD-TYPE.
FT   NP_BIND    1594   1601       ATP (POTENTIAL).
FT   SITE       1719   1722       DEGH BOX.
FT   DOMAIN      745    750       POLY-SER.
FT   DOMAIN     1151   1156       POLY-SER.
FT   DOMAIN     1166   1169       POLY-LYS.
FT   DOMAIN     1202   1206       POLY-SER.
FT   DOMAIN     1259   1266       POLY-ASP.
FT   DOMAIN     1443   1466       POLY-GLU.
FT   DOMAIN     1499   1502       POLY-GLU.
FT   DOMAIN     1929   1939       POLY-LYS.
FT   DOMAIN     1941   1948       POLY-SER.
FT   DOMAIN     2222   2225       POLY-LYS.
FT   DOMAIN     2262   2265       POLY-GLU.
FT   DOMAIN     2420   2425.      POLY-GLN.
SQ   SEQUENCE   2492 AA;  282567 MW;  56978AC7D37400DC CRC64;

Alignment Scores:
Pred. No.:              0.00499       Length:        2492
Score:                  162.00        Matches:       175
Percent Similarity:     36.09%        Conservative:  135
Best Local Similarity:  20.37%        Mismatches:    314
Query Match:            2.25%         Indels:        235
DB:                     1             Gaps:          47

US-09-302-812-3 (1-4069) x ATRX_PANTR (1-2492)
```

```
Qy        389 AAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448
              :::::::::   ::::::|||         :::||||||   |||::::::      |||
Db       1136 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSerSer 1155

Qy        449 AGTTTGGATAGTAAAGAAAAC------AACAATACAAGAATAGAATCCATGATGAGTTCT 502
              |||   |||::::::|||:::         :::|||   :::   :::      ||||||
Db       1156 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysGlnGlnArgThrSerSer 1174

Qy        503 GTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA 562
                 :::|||         :::                     :::      |||
Db       1175 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 1194

Qy        563 AGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCA 622
```

```
Db  1195 ArgLysGlnAlaAspIleThrSerSerSerSerSerAspIleGluAspAspAspGlnAsn 1214
Qy   623 GCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCT 682
Db  1215 SerIle------------GlyGluGlySerSerAspGluGlnLysIle------LysPro 1228
Qy   683 CAAACAGTAACCCTGGTACCAGAGCAG---------TTTAGTAATGCTAACATTGATCGG 733
Db  1229 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 1248
Qy   734 TCA---------------------CCTCAAAATGATGATCACAGTGACACAGATAGTGAA 772
Db  1249 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAsnAspProGlu 1268
Qy   773 GAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------CTTGCAAATGCAAAG 826
Db  1269 AsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuSerSerAspGlu 1288
Qy   827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
Db  1289 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 1308
Qy   887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
Db  1309 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer------------------ 1322
Qy   947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGAC------ 1000
Db  1323 ------GluSerAspSerAspSerGluGlu------------SerLysLysProArgTyr 1336
Qy  1001 ---AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCA---TTTGAGAAG 1054
Db  1337 ArgHisArgLeuLeuArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLys 1356
Qy  1055 GAAAGTGAACCCGAATCACCGATGGATGTG---------GATAATTCTAAAAATAGTTGT 1105
Db  1357 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 1376
Qy  1106 CAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGT 1165
Db  1377 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal------Ser 1394
Qy  1166 TCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTT 1225
Db  1395 GluSerGluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGlu 1414
Qy  1226 AGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGA 1282
Db  1415 AsnGlnArgSerTyrLysGlnLysLysLysArgArgArgIleLysValGlnGluAspSer 1434
Qy  1283 GGA--------------------------------------------------------- 1285
Db  1435 SerSerGluAsnLysSerAsnSerGluGluGluGluGluGluLysGluGluGluGluGlu 1454
Qy  1286 ------------------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCT 1327
Db  1455 GluGluGluGluGluGluGluGluGluGluAspGluAsnAsp---AspSerLysSerPro 1473
Qy  1328 GGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAA------ 1381
Db  1474 Gly---------------------------LysGlyArgLysLysIleArgLysIleLeu 1484
Qy  1382 AAGGATTCTAAAATC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
Db  1485 LysAspAspLysLeuArgThrGluThrGlnAsnAlaLeuLysGluGluGluGluArgArg 1504
Qy  1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA 1498
Db  1505 LysArgIleAlaGluArgGluArgGluArg------GluLysLeuArgGluValIleGlu 1522
Qy  1499 ---CCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT--------------- 1540
Db  1523 IleGluAspAlaSerProThrLysCysProIleThrThrLysLeuValLeuAspGluAsp 1542
Qy  1541 GAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCC 1600
Db  1543 GluGluThrLysGlu------------------------ProLeuValGlnValHisArg 1554
Qy  1601 AATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTT 1660
Db  1555 AsnMetValIleLysLeu---------------------------------LysPro--- 1562
Qy  1661 CCAACACATTATAAAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCT 1708
Db  1563 ------HisGlnValAspGlyValGlnPheMetTrpAsp--------------------- 1573
Qy  1709 TGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGC 1768
Db  1574 CysCysCysGluSerVal---------------LysLysThrLysLysSerProGlySer 1588
Qy  1769 CGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAG 1828
Db  1589 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys---------------ThrLeuGln 1603
Qy  1829 GATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTG 1885
Db  1604 ValValSerPheLeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeu 1623
Qy  1886 ATC------------------------------------------GATTTCTGGGATAAG 1903
Db  1624 ValValCysProLeuAsnThrAlaLeuAsnTrpMetAsnGluPheGluLysTrpGlnGlu 1643
Qy  1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db  1644 GlyLeuLysAspAspGlu------------------------------------------ 1649
Qy  1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db  1650 ---LysLeuGluValSerGluLeuAlaThrValLysArgPro---------GlnGluArg 1665
Qy  2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db  1666 SerTyrMetLeuGlnArgTrpGlnGluAspGlyGlyValMetIleIleGly--------- 1682
Qy  2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db  1683 ------------------------------------TyrGluMetTyr------------ 1686
Qy  2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db  1687 ------ArgAsnLeuAlaGlnGlyArgAsnVal---LysSerArgLysLeuLysGluIle 1703
Qy  2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db  1704 Phe------AsnLysAlaLeuValAspProGlyProAspPheValValCysAspGluGly 1721
Qy  2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------------ 2311
Db  1722 HisIleLeuLysAsn------GluAlaSerAlaValSerLysAlaMetAsnSerIleArg 1739
Qy  2312 ACACGATTGCATGTCACTTACGAAGGTACC------------------------------ 2341
Db  1740 SerArgArgArgIleIleLeuThrGlyThrProLeuGlnAsnAsnLeuIleGluTyrHis 1759
Qy  2342 ---------------ATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db  1760 CysMetValAsnPheIleLysGluAsnLeuLeuGlySerIleLys---GluPheArgAsn 1778
Qy  2387 CGTTTTGTT------------GGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAA 2434
Db  1779 ArgPheIleAsnProIleGlnAsnGlyGlnCysAlaAspSerThrMetVal------Asp 1796
Qy  2435 ATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTC------------------ 2476
Db  1797 ValArgValMetLysLysArgAlaHisIleLeuTyrGluMetLeuAlaGlyCysValGln 1816
```

```
Qy       2477 ------------------TTCACTGAGGTGCTG------GATCACAATGAATGTCTAATT 2512
                                  |||:::   |||         |||        |||
Db       1817 ArgLysAspTyrThrAlaLeuThrLysPheLeuProProLysHisGluTyrValLeuAla 1836

Qy       2513 ATCACAGGTACTGAG---CAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGG 2569
               :::       |||      |||          |||      ||| |||    |||:::   ||
Db       1837 ValArgMetThrSerIleGlnCysLysLeuTyrGlnTyrTyr-LeuAspHisLeuThrGl 1856

Qy       2570 TCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTG 2624
               |   |||             |||   :::|||         :::||||||    |||
Db       1856 yValGlyAsnAsnSerGluGlyGlyArgGly------LysAlaGlyAlaLysLeu 1872

RESULT 7
ATRX_HUMAN
ID   ATRX_HUMAN     STANDARD;      PRT;  2492 AA.
AC   P46100; P51068; Q15886; Q9H0Z1; Q9NTS3;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP) (Znf-HX).
GN   ATRX OR RAD54L OR XH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;
RP   GLU-740 AND SER-1860, AND VARIANTS ATR-X.
RX   MEDLINE=97123494; PubMed=8968741;
RA   Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA   Gibbons R.J.;
RT   "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT   point to a common mechanism underlying the ATR-X syndrome.";
RL   Hum. Mol. Genet. 5:1899-1907(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX   MEDLINE=97386582; PubMed=9244431;
RA   Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA   Colleaux L., Schwartz C., Fontes M.;
RT   "Determination of the genomic structure of the XNP/ATRX gene encoding
RT   a potential zinc finger helicase.";
RL   Genomics 43:149-155(1997).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS SER-596 AND GLU-740.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
RN   [4]
RP   SEQUENCE OF 860-2492 FROM N.A.
RX   MEDLINE=95179111; PubMed=7874112;
RA   Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RA   Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA   Bianchi M.E., Consalez G.G.;
RT   "Cloning and characterization of a new human Xq13 gene, encoding a
RT   putative helicase.";
RL   Hum. Mol. Genet. 3:1957-1964(1994).
RN   [5]
RP   PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94214473; PubMed=8162050;
RA   Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA   Millasseau P., Khrestchatisky M., Fontes M.;
RT   "Cloning and expression of the murine homologue of a putative human
RT   X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL   Hum. Mol. Genet. 3:39-44(1994).
RN   [6]
RP   SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX   MEDLINE=95211835; PubMed=7697714;
RA   Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT   "Mutations in a putative global transcriptional regulator cause X-
RT   linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL   Cell 80:837-845(1995).
RN   [7]
RP   SEQUENCE OF 1375-2492 FROM N.A.
RA   Pearce A., Chapman J.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   EZH2 BINDING.
RX   MEDLINE=98167853; PubMed=9499421;
RA   Cardoso C., Timsit S., Villard L., Khrestchatisky M., Fontes M.,
RA   Colleaux L.;
RT   "Specific interaction between the XNP/ATR-X gene product and the SET
RT   domain of the human EZH2 protein.";
RL   Hum. Mol. Genet. 7:679-684(1998).
RN   [9]
RP   SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP   HETEROCHROMATIN.
RX   MEDLINE=20040663; PubMed=10570185;
RA   McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA   Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA   Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT   "Localization of a putative transcriptional regulator (ATRX) at
RT   pericentromeric heterochromatin and the short arms of acrocentric
RT   chromosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN   [10]
RP   DISEASE.
RX   MEDLINE=20213147; PubMed=10751095;
RA   Villard L., Fontes M., Ades L.C., Gecz J.;
RT   "Identification of a mutation in the XNP/ATR-X gene in a family
RT   reported as Smith-Fineman-Myers syndrome.";
RL   Am. J. Med. Genet. 91:83-85(2000).
RN   [11]
RP   DISEASE.
RX   MEDLINE=22804448; PubMed=12858175;
RA   Gibbons R.J., Pellagatti A., Garrick D., Wood W.G., Malik N.,
RA   Ayyub H., Langford C., Boultwood J., Wainscoat J.S., Higgs D.R.;
RT   "Identification of acquired somatic mutations in the gene encoding
RT   chromatin-remodeling factor ATRX in the alpha-thalassemia
RT   myelodysplasia syndrome (ATMDS).";
RL   Nat. Genet. 34:446-449(2003).
RN   [12]
RP   VARIANT ATR-X SER-1713.
RX   MEDLINE=97196774; PubMed=9043863;
RA   Villard L., Lacombe D., Fontes M.;
RT   "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT   without alpha-thalassemia.";
RL   Eur. J. Hum. Genet. 4:316-320(1996).
RN   [13]
RP   VARIANT JM GLN-2131.
RX   MEDLINE=96224392; PubMed=8630485;
RA   Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Veber P.,
RA   Munnich A., Lyonnet S.;
RT   "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL   Nat. Genet. 12:359-360(1996).
RN   [14]
RP   VARIANTS ATR-X.
RX   MEDLINE=97467722; PubMed=9326931;
RA   Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA   Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA   Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA   Higgs D.R.;
RT   "Mutations in transcriptional regulator ATRX establish the functional
RT   significance of a PHD-like domain.";
RL   Nat. Genet. 17:146-148(1997).
RN   [15]
RP   VARIANT ATR-X LEU-246.
RX   MEDLINE=20123062; PubMed=10660327;
RA   Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA   Greco D., Cardoso C., Fontes M., Ragusa A.;
RT   "New mutations in XNP/ATR-X gene: a further contribution to
RT   genotype/phenotype relationship in ATR/X syndrome.";
RL   Hum. Mutat. 12:214-214(1998).
```

```
RN   [16]
RP   VARIANT SHS LYS-1742.
RX   MEDLINE=99347960; PubMed=10417298;
RA   Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA   Prieto F., Fontes M., Martinez F.;
RT   "Mutation of the XNP/ATR-X gene in a family with severe mental
RT   retardation, spastic paraplegia and skewed pattern of X inactivation:
RT   demonstration that the mutation is involved in the inactivation
RT   bias.";
RL   Am. J. Hum. Genet. 65:558-562(1999).
RN   [17]
RP   VARIANT CWS THR-2050.
RX   MEDLINE=99326061; PubMed=10398237;
RA   Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
RA   Curtis M.;
RT   "Carpenter-Waziri syndrome results from a mutation in XNP.";
RL   Am. J. Med. Genet. 85:249-251(1999).
RN   [18]
RP   VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;
RP   LEU-246 AND CYS-249.
RX   MEDLINE=99219535; PubMed=10204841;
RA   Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougne J.,
RA   Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA   Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
RT   "Evaluation of a mutation screening strategy for sporadic cases of
RT   ATR-X syndrome.";
RL   J. Med. Genet. 36:183-186(1999).
RN   [19]
RP   VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1552; SER-1645
RP   AND CYS-1847.
RX   MEDLINE=20451413; PubMed=10995512;
RA   Wada T., Kubota T., Fukushima Y., Saitoh S.;
RT   "Molecular genetic study of Japanese patients with X-linked alpha-
RT   thalassemia/mental retardation syndrome (ATR-X).";
RL   Am. J. Med. Genet. 94:242-248(2000).
RN   [20]
RP   VARIANT ATR-X MET-1621.
RX   MEDLINE=22108790; PubMed=12116232;
RA   Yntema H.G., Poppelaars F.A., Derksen E., Oudakker A.R.,
RA   van Roosmalen T., Jacobs A., Obbema H., Brunner H.G., Hamel B.C.J.,
RA   van Bokhoven H.;
RT   "Expanding phenotype of XNP mutations: mild to moderate mental
RT   retardation.";
RL   Am. J. Med. Genet. 110:243-247(2002).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC       Name=4;
CC         IsoId=P46100-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=P46100-2; Sequence=VSP_000575;
CC       Name=2;
CC         IsoId=P46100-3; Sequence=VSP_000574;
CC       Name=3;
CC         IsoId=P46100-4; Sequence=VSP_000576;
CC       Name=5;
CC         IsoId=P46100-5; Sequence=VSP_000574, VSP_000576;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DISEASE: Defects in ATRX are the cause of X-linked alpha-
CC       thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-
CC       X is an X-linked disorder comprising severe psychomotor

Alignment Scores:
Pred. No.:          0.00582      Length:      2492
Score:              161.00       Matches:     175
```

```
Percent Similarity:     36.09%      Conservative:   135
Best Local Similarity:  20.37%      Mismatches:     314
Query Match:            2.23%       Indels:         235
DB:                     1           Gaps:           47

US-09-302-812-3 (1-4069) x ATRX_HUMAN (1-2492)

Qy     389 AAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448
           ::::::::   ::::::|||       :::||||||   |||::::::       |||
Db    1136 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSerSer 1155

Qy     449 AGTTTGGATAGTAAAGAAAAC------AACAATACAAGAATAGAATCCATGATGAGTTCT 502
            |||   |||::::::|||:::       :::|||    :::    :::      ||||||
Db    1156 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysLysGlnArgThrSerSer 1174

Qy     503 GTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA 562
              :::|||            :::                  :::      |||
Db    1175 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 1194

Qy     563 AGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCA 622
                             :::|||       |||:::      :::
Db    1195 ArgLysGlnAlaAspIleThrSerSerSerSerSerAspIleGluAspAspAspGlnAsn 1214

Qy     623 GCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTGGAAAGTGAACCT 682
           ::::::                  ||||||      ||||||    :::        :::|||
Db    1215 SerIle------------GlyGluGlySerSerAspGluGlnLysIle------LysPro 1228

Qy     683 CAAACAGTAACCCTGGTACCAGAGCAG---------TTTAGTAATGCTAACATTGATCGG 733
              |||       ||||||                  |||         ::::::  |||
Db    1229 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 1248

Qy     734 TCA---------------------CCTCAAAATGATGATCACAGTGACACAGATAGTGAA 772
           :::                        :::||||||      |||      |||    |||
Db    1249 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAsnAspProGlu 1268

Qy     773 GAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------CTTGCAAATGCAAAG 826
                             :::      |||      :::|||      |||::::::   :::
Db    1269 AsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuSerSerAspGlu 1288

Qy     827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
                 :::      |||:::            |||      |||
Db    1289 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 1308

Qy     887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
           :::||||||::::::      |||      |||      :::
Db    1309 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer------------------ 1322

Qy     947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGAC------ 1000
                    ||||||      ||||||:::            ||||||
Db    1323 ------GluSerAspSerAspSerGluGlu------------SerLysLysProArgTyr 1336

Qy    1001 ---AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCA---TTTGAGAAG 1054
              ::::::|||:::|||      :::      :::      :::            ||||||
Db    1337 ArgHisArgLeuLeuArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLys 1356

Qy    1055 GAAAGTGAACCCGAATCACCGATGGATGTG---------GATAATTCTAAAAATAGTTGT 1105
           :::::::::|||:::            :::|||      :::      |||   |||
Db    1357 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 1376

Qy    1106 CAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGT 1165
           :::|||||||||      :::            |||      |||:::            |||
Db    1377 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal------Ser 1394

Qy    1166 TCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTT 1225
              |||:::            :::|||            |||:::   |||::::::|||
Db    1395 GluSerGluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGlu 1414

Qy    1226 AGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGA 1282
              :::|||   |||         |||            |||:::      |||   :::
Db    1415 AsnGlnArgSerTyrLysGlnLysLysLysArgArgArgIleLysValGlnGluAspSer 1434

Qy    1283 GGA--------------------------------------------------------- 1285
```

```
Db    1435 SerSerGluAsnLysSerAsnSerGluGluGluGluGluLysGluGluGluGluGlu 1454

Qy    1286 -----------------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCT 1327
                            |||              ||||||   :::::|||   |||
Db    1455 GluGluGluGluGluGluGluGluGluGluAspGluAsnAsp---AspSerLysSerPro 1473

Qy    1328 GGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAA------ 1381
           |||                          :::         |||:::      |||
Db    1474 Gly--------------------------LysGlyArgLysLysIleArgLysIleLeu 1484

Qy    1382 AAGGATTCTAAAATC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
           ||||||  |||:::   |||:::              |||     :::   |||:::||||||
Db    1485 LysAspAspLysLeuArgThrGluThrGlnAsnAlaLeuLysGluGluGluGluArgArg 1504

Qy    1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA 1498
           |||         |||   :::   :::|||          |||:::   :::   :::
Db    1505 LysArgIleAlaGluArgGluArgGluArg------GluLysLeuArgGluValIleGlu 1522

Qy    1499 ---CCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT--------------- 1540
                     ||||||  |||      :::   |||    :::
Db    1523 IleGluAspAlaSerProThrLysCysProIleThrThrLysLeuValLeuAspGluAsp 1542

Qy    1541 GAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCC 1600
           ||||||   :::                       ||||||::::::
Db    1543 GluGluThrLysGlu-----------------------ProLeuValGlnValHisArg 1554

Qy    1601 AATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTT 1660
           |||      :::   :::                                ||||||
Db    1555 AsnMetValIleLysLeu--------------------------------LysPro--- 1562

Qy    1661 CCAACACATTATAAAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCT 1708
                 |||     |||               :::||||||
Db    1563 ------HisGlnValAspGlyValGlnPheMetTrpAsp-------------------- 1573

Qy    1709 TGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGC 1768
           |||       :::::::::              :::      :::::::::  ||||||
Db    1574 CysCysCysGluSerVal---------------LysLysThrLysLysSerProGlySer 1588

Qy    1769 CGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAG 1828
                    |||              |||   |||                |||:::
Db    1589 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys---------------ThrLeuGln 1603

Qy    1829 GATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTG 1885
                       :::   |||            |||   ||||||   |||||||||
Db    1604 ValValSerPheLeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeu 1623

Qy    1886 ATC------------------------------------------GATTTCTGGGATAAG 1903
           :::                                          :::   |||   :::
Db    1624 ValValCysProLeuAsnThrAlaLeuAsnTrpMetAsnGluPheGluLysTrpGlnGlu 1643

Qy    1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              |||::::::   |||
Db    1644 GlyLeuLysAspAspGlu------------------------------------------ 1649

Qy    1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
                 |||   :::       :::   |||          |||         |||:::
Db    1650 ---LysLeuGluValSerGluLeuAlaThrValLysArgPro---------GlnGluArg 1665

Qy    2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
           ::::::   :::            ||||||          ::::::
Db    1666 SerTyrMetLeuGlnArgTrpGlnGluAspGlyGlyValMetIleIleGly--------- 1682

Qy    2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
                                                    |||      |||
Db    1683 ---------------------------------------TyrGluMetTyr--------- 1686

Qy    2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
                    |||   :::||||||:::      |||         |||||||||   :::
Db    1687 ------ArgAsnLeuAlaGlnGlyArgAsnVal---LysSerArgLysLeuLysGluIle 1703

Qy    2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
           |||         :::   :::   :::         |||      :::|||
Db    1704 Phe------AsnLysAlaLeuValAspProGlyProAspPheValValCysAspGluGly 1721

Qy    2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG----------- 2311
              |||::::::          |||                  |||   :::
Db    1722 HisIleLeuLysAsn------GluAlaSerAlaValSerLysAlaMetAsnSerIleArg 1739

Qy    2312 ACACGATTGCATGTCACTTACGAAGGTACC------------------------------ 2341
           :::|||      :::          ||||||
Db    1740 SerArgArgArgIleIleLeuThrGlyThrProLeuGlnAsnAsnLeuIleGluTyrHis 1759

Qy    2342 --------------ATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                         |||:::||||||      |||   ::::::   :::|||   |||
Db    1760 CysMetValAsnPheIleLysGluAsnLeuLeuGlySerIleLys---GluPheArgAsn 1778

Qy    2387 CGTTTTGTT-----------GGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAA 2434
           ||||||:::               |||          :::   :::|||         :::
Db    1779 ArgPheIleAsnProIleGlnAsnGlyGlnCysAlaAspSerThrMetVal------Asp 1796

Qy    2435 ATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTC------------------ 2476
           :::|||   :::                 |||:::      :::
Db    1797 ValArgValMetLysLysArgAlaHisIleLeuTyrGluMetLeuAlaGlyCysValGln 1816

Qy    2477 -----------------TTCACTGAGGTGCTG------GATCACAATGAATGTCTAATT 2512
                            |||:::   |||            |||            |||
Db    1817 ArgLysAspTyrThrAlaLeuThrLysPheLeuProProLysHisGluTyrValLeuAla 1836

Qy    2513 ATCACAGGTACTGAG---CAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGG 2569
           :::      |||      |||         |||      |||  |||      |||:::   ||
Db    1837 ValArgMetThrSerIleGlnCysLysLeuTyrGlnTyrTyr-LeuAspHisLeuThrGl 1856

Qy    2570 TCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTG 2624
            |    |||          |||   :::|||         :::||||||   |||
Db    1856 yValGlyAsnAsnSerGluGlyGlyArgGly------LysAlaGlyAlaLysLeu 1872
```

```
RESULT 8
ATRX_PONPY
ID   ATRX_PONPY     STANDARD;      PRT;  2492 AA.
AC   Q7YQM3;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE   nuclear protein) (XNP).
GN   ATRX.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT   chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
CC   -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC       gene expression by affecting chromatin. May be involved in brain
CC       development and facial morphogenesis.
CC   -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC       phosphatidylcholine/phosphatidylserine-dependent manner (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC       heterochromatin during interphase and mitosis, probably by
CC       interacting with HP1.
CC   -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; AB102643; BAC81112.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW   Zinc-finger.
FT   ZN_FING       220    268       PHD-TYPE.
FT   NP_BIND      1594   1601       ATP (POTENTIAL).
FT   SITE         1719   1722       DEGH BOX.
FT   DOMAIN        745    750       POLY-SER.
FT   DOMAIN       1151   1156       POLY-SER.
FT   DOMAIN       1166   1169       POLY-LYS.
FT   DOMAIN       1202   1206       POLY-SER.
FT   DOMAIN       1259   1266       POLY-ASP.
FT   DOMAIN       1443   1466       POLY-GLU.
FT   DOMAIN       1499   1502       POLY-GLU.
FT   DOMAIN       1929   1939       POLY-LYS.
FT   DOMAIN       1941   1948       POLY-SER.
FT   DOMAIN       2222   2225       POLY-LYS.
FT   DOMAIN       2262   2265       POLY-GLU.
FT   DOMAIN       2420   2425       POLY-GLN.
SQ   SEQUENCE   2492 AA;  282614 MW;  48EC97106D0CEF20 CRC64;

Alignment Scores:
Pred. No.:                0.00582        Length:        2492
Score:                    161.00         Matches:       175
Percent Similarity:       36.09%         Conservative:  135
Best Local Similarity:    20.37%         Mismatches:    314
Query Match:              2.23%          Indels:        235
DB:                       1              Gaps:          47

US-09-302-812-3 (1-4069) x ATRX_PONPY (1-2492)

Qy        389 AAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448
               :::::::::   ::::::|||       :::||||||   |||::::::        |||
Db       1136 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSerSer 1155

Qy        449 AGTTTGGATAGTAAAGAAAAC------AACAATACAAGAATAGAATCCATGATGAGTTCT 502
               |||   |||::::::|||:::      :::|||   :::   :::        ||||||
Db       1156 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysLysGlnArgThrSerSer 1174

Qy        503 GTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA 562
                 :::|||            :::                :::        |||
Db       1175 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 1194

Qy        563 AGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCA 622
                 :::|||         |||:::        :::
Db       1195 ArgLysGlnAlaAspIleThrSerSerSerSerSerAspIleGluAspAspAspGlnAsn 1214

Qy        623 GCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCT 682
              ::::::            ||||||         ||||||   :::      :::|||
Db       1215 SerIle------------GlyGluGlySerSerAspGluGlnLysIle------LysPro 1228

Qy        683 CAAACAGTAACCCTGGTACCAGAGCAG---------TTTAGTAATGCTAACATTGATCGG 733
                 |||      ||||||            |||      ::::::   |||
Db       1229 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 1248

Qy        734 TCA--------------------CCTCAAAATGATGATCACAGTGACACAGATAGTGAA 772
              :::                     :::||||||      |||   |||   |||
Db       1249 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAsnAspProGlu 1268

Qy        773 GAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG------CTTGCAAATGCAAAG 826
                       :::   |||      :::|||   |||::::::   :::
Db       1269 AsnArgIleAlaLysLysMetLeuLeuGluGluIleLysAlaAsnLeuSerSerAspGlu 1288

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
                 :::   |||:::            :::|||   |||
Db       1289 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 1308

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              :::||||||::::::   |||         |||   :::
Db       1309 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer------------------ 1322

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGAC------ 1000
                    ||||||      ||||||:::            ||||||
Db       1323 ------GluSerAspSerAspSerGluGlu------------SerLysLysProArgTyr 1336

Qy       1001 ---AACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCA---TTTGAGAAG 1054
              ::::::|||:::|||   :::      :::   :::               ||||||
Db       1337 ArgHisArgLeuLeuArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLys 1356

Qy       1055 GAAAGTGAACCCGAATCACCGATGGATGTG---------GATAATTCTAAAAATAGTTGT 1105
              :::::::::|||:::         :::|||         :::      |||   |||
Db       1357 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 1376

Qy       1106 CAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGT 1165
              :::|||||||||      :::            |||      |||:::         |||
Db       1377 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluVal------Ser 1394

Qy       1166 TCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTT 1225
                 |||:::      :::|||      |||:::      |||::::::|||
Db       1395 GluSerGluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGlu 1414

Qy       1226 AGGAAACGG---TACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGA 1282
                 :::|||   |||      |||         |||:::      |||   :::
Db       1415 AsnGlnArgSerTyrLysGlnLysLysLysArgArgArgIleLysValGlnGluAspSer 1434

Qy       1283 GGA--------------------------------------------------------- 1285

Db       1435 SerSerGluAsnLysSerAsnSerGluGluGluGluGluLysGluGluGluGluGluGlu 1454

Qy       1286 ------------------GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCT 1327
                                |||         ||||||   ::::::|||   |||
Db       1455 GluGluGluGluGluGluGluGluGluGluAspGluAsnAsp---AspSerLysSerPro 1473

Qy       1328 GGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAA------ 1381
              |||                        :::      |||:::      |||
Db       1474 Gly-------------------------LysGlyArgLysLysIleArgLysIleLeu 1484

Qy       1382 AAGGATTCTAAAATC---ACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
              ||||||   |||:::   |||:::         |||   :::   |||:::||||||
Db       1485 LysAspAspLysLeuArgThrGluThrGlnAsnAlaLeuLysGluGluGluGluArgArg 1504

Qy       1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCA 1498
              |||         |||   :::   :::|||         |||:::   :::   :::
Db       1505 LysArgIleAlaGluArgGluArgGluArg------GluLysLeuArgGluValIleGlu 1522

Qy       1499 ---CCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT--------------- 1540
                    ||||||   |||         :::   |||   :::
Db       1523 IleGluAspAlaSerProThrLysCysProIleThrThrLysLeuValLeuAspGluAsp 1542

Qy       1541 GAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCC 1600
              ||||||   :::                        ||||||::::::
Db       1543 GluGluThrLysGlu-----------------------ProLeuValGlnValHisArg 1554

Qy       1601 AATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTT 1660
              |||      :::   :::                                   ||||||
Db       1555 AsnMetValIleLysLeu---------------------------------LysPro--- 1562

Qy       1661 CCAACACATTATAAAGAT------------TTGTGGGATAACAAGCATGTTAAAATGCCT 1708
                    |||      |||         :::||||||
Db       1563 ------HisGlnValAspGlyValGlnPheMetTrpAsp--------------------- 1573

Qy       1709 TGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGC 1768
              |||      :::::::::               :::      :::::::::   ||||||
Db       1574 CysCysCysGluSerVal---------------LysLysThrLysLysSerProGlySer 1588

Qy       1769 CGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAG 1828
                    |||         |||   |||                     |||:::
Db       1589 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys---------------ThrLeuGln 1603
```

```
Qy       1829 GATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTT---ACAGCTTTG 1885
                        :::  |||               |||  ||||||   |||||||||
Db       1604 ValValSerPheLeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeu 1623

Qy       1886 ATC------------------------------------------GATTTCTGGGATAAG 1903
              :::                                          :::  |||  :::
Db       1624 ValValCysProLeuAsnThrAlaLeuAsnTrpMetAsnGluPheGluLysTrpGlnGlu 1643

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
                 |||::::::  |||
Db       1644 GlyLeuLysAspAspGlu------------------------------------------ 1649

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
                 |||  :::    :::  |||         |||         |||:::
Db       1650 ---LysLeuGluValSerGluLeuAlaThrValLysArgPro---------GlnGluArg 1665

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              ::::::  :::        ||||||        ::::::
Db       1666 SerTyrMetLeuGlnArgTrpGlnGluAspGlyGlyValMetIleIleGly--------- 1682

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
                                                     |||   |||
Db       1683 ---------------------------------------TyrGluMetTyr--------- 1686

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
                       |||  :::||||||:::     |||      |||||||||  :::
Db       1687 ------ArgAsnLeuAlaGlnGlyArgAsnVal---LysSerArgLysLeuLysGluIle 1703

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              |||        :::  :::  :::        |||     :::|||
Db       1704 Phe------AsnLysAlaLeuValAspProGlyProAspPheValValCysAspGluGly 1721

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------------ 2311
                 |||::::::        |||         |||  :::
Db       1722 HisIleLeuLysAsn------GluAlaSerAlaValSerLysAlaMetAsnSerIleArg 1739

Qy       2312 ACACGATTGCATGTCACTTACGAAGGTACC------------------------------ 2341
              :::|||      :::        ||||||
Db       1740 SerArgArgArgIleIleLeuThrGlyThrProLeuGlnAsnAsnLeuIleGluTyrHis 1759

Qy       2342 ---------------ATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                             |||:::||||||     |||   ::::::   :::|||   |||
Db       1760 CysMetValAsnPheIleLysGluAsnLeuLeuGlySerIleLys---GluPheArgAsn 1778

Qy       2387 CGTTTTGTT-----------GGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAA 2434
              ||||||:::                 |||        :::  :::|||         :::
Db       1779 ArgPheIleAsnProIleGlnAsnGlyGlnCysAlaAspSerThrMetVal------Asp 1796

Qy       2435 ATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTC------------------ 2476
              :::|||    :::              |||:::     :::
Db       1797 ValArgValMetLysLysArgAlaHisIleLeuTyrGluMetLeuAlaGlyCysValGln 1816

Qy       2477 -----------------TTCACTGAGGTGCTG------GATCACAATGAATGTCTAATT 2512
                                |||:::     |||         |||         |||
Db       1817 ArgLysAspTyrThrAlaLeuThrLysPheLeuProProLysHisGluTyrValLeuAla 1836

Qy       2513 ATCACAGGTACTGAG---CAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGG 2569
              :::       |||     |||           |||      ||| |||    |||:::   ||
Db       1837 ValArgMetThrSerIleGlnCysLysLeuTyrGlnTyrTyr-LeuAspHisLeuThrGl 1856

Qy       2570 TCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTG 2624
               \    |||             |||   :::|||           :::||||||   |||
Db       1856 yValGlyAsnAsnSerGluGlyGlyArgGly------LysAlaGlyAlaLysLeu 1872
```

```
RESULT 9
SPOF_SCHPO
ID   SPOF_SCHPO     STANDARD;      PRT;  1957 AA.
AC   Q10411; Q9USE9;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Sporulation-specific protein 15.
GN   SPO15 OR SPAC1F3.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=20107136; PubMed=10639340;
RA   Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT   "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
RT   to the spindle pole body and essential for its modification.";
RL   J. Cell Sci. 113:545-554(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   SEQUENCE OF 705-871 FROM N.A.
RC   STRAIN=968 h90;
RX   MEDLINE=20223868; PubMed=10759889;
RA   Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA   Hiraoka Y.;
RT   "Large-scale screening of intracellular protein localization in living
RT   fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL   Genes Cells 5:169-190(2000).
CC   -!- FUNCTION: Has a role in the initiation of spore membrane
CC       formation.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Spindle pole body.
CC   -!- SIMILARITY: Belongs to the MPC70 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z70690; CAA94624.1; -.
DR   EMBL; AB027811; BAA87115.1; -.
DR   PIR; T38077; T38077.
DR   GeneDB_SPombe; SPAC1F3.06c; -.
KW   Sporulation; Coiled coil.
FT   DOMAIN      199    785       COILED COIL (POTENTIAL).
FT   DOMAIN      804   1235       COILED COIL (POTENTIAL).
FT   DOMAIN     1320   1471       COILED COIL (POTENTIAL).
```

```
FT   DOMAIN      1481   1723       COILED COIL (POTENTIAL).
SQ   SEQUENCE    1957 AA;  222785 MW;  3F480CA06171D9DA CRC64;

Alignment Scores:
Pred. No.:              0.00637        Length:        1957
Score:                  160.00         Matches:       182
Percent Similarity:     35.63%         Conservative:  149
Best Local Similarity:  19.59%         Mismatches:    343
Query Match:            2.22%          Indels:        255
DB:                     1              Gaps:          39

US-09-302-812-3 (1-4069) x SPOF_SCHPO (1-1957)

Qy        242 TCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCAAGGACGCTCAC 301
              |||         |||       |||   ||||||    :::|||:::    |||      |||
Db        508 SerThrGlnSerSerLeuGlnSerLeuGlnArgAspIleLeuAsnGluLysLysLysHis 527

Qy        302 GTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGGCGGGACAGCAC 361
                    :::                                         |||     :::
Db        528 GluValTyrGlu-----------------------------SerGlnLeuAsnGluLeu 537

Qy        362 AGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACT 421
              :::|||         :::   :::           :::  ::::::|||  :::    |||
Db        538 LysGlyGluLeuGlnThrGluIleSerAsnSerGluHisLeuSerSerGlnLeuSerThr 557

Qy        422 AAAGGAATCAAGACACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAACAATACAAGA 481
                                 ||||||    |||:::        :::          ||||||
Db        558 ---------LeuAlaAlaGluLysGluAla------AlaValAlaThrAsnAsnGluLeu 572

Qy        482 ATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAACATAATGTAGAAAAA 541
                 ||||||      |||                   |||   :::|||       :::   |||
Db        573 SerGluSerLysAsnSerLeuGlnThrLeuCysAsnAlaPheGlnGluLysLeuAlaLys 592

Qy        542 TTAGTAAATGTTTCTCAGCTA------------------------AGTCTTGAT------ 571
                    :::|||    ||||||                         |||||||||
Db        593 ------SerValMetGlnLeuLysGluAsnGluGlnAsnPheSerSerLeuAspThrSer 610

Qy        572 ---AAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATG 628
                 |||    |||   |||               :::       |||    |||||||||
Db        611 PheLysLysLeuAsnGluSerHisGlnGluLeuGluAsnAsnHisGlnThrIle------ 628

Qy        629 TGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACA 688
                                           |||:::||||||     ::::::      :::
Db        629 --------------------------ThrLysGlnLeuLysAspThrSerSerLysLeu 639

Qy        689 GTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGAT 748
                     |||          |||         ||||||   ::::::
Db        640 GlnGlnLeu---------GlnLeuGluArgAlaAsnPheGluGlnLys------------ 652

Qy        749 GATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTA 808
                              |||   |||:::||||||    |||      :::        |||          :::
Db        653 ------GluSerThrLeuSerAspGluAsnAsnAspLeuArgThrLysLeuLeuLysLeu 670

Qy        809 AAGCTTGCAAATGCAAAG------------------------------------------ 826
              :::    :::|||
Db        671 GluGluSerAsnLysSerLeuIleLysLysGlnGluAspValAspSerLeuGluLysAsn 690

Qy        827 ---CAGACTACGGAAGATGAACACGCCAGA------------------------------ 853
                 ||||||   :::::::::       :::
Db        691 IleGlnThrLeuLysGluAspLeuArgLysSerGluGluAlaLeuArgPheSerLysLeu 710

Qy        854 GAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGT 913
              |||||||||:::    ::::::           :::                :::
Db        711 GluAlaLysAsnLeuArgGluValIleAspAsnLeuLysGlyLysHisGluThrLeuGlu 730

Qy        914 TGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAG 973
                 |||::::::::::::                   |||    |||||||||
Db        731 AlaGlnArgAsnAspLeu------------HisSerSerLeuSerAsp------------ 742

Qy        974 GATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGA 1033
                                 :::||||||    |||    ::::::          |||
Db        743 ---------------AlaLysAsnThrAsnAlaIleLeuSerSerGlu------------ 753

Qy       1034 AATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCT 1093
                          |||  ||||||                 :::         |||
Db        754 ------------LeuThrLysSerSerGluAspValLysArgLeuThrAla---AsnVal 768

Qy       1094 AAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAA 1153
              :::             |||||||||:::|||  |||::::::                 |||:::|||
Db        769 GluThrLeuThrGlnAspSerLysAla-MetLysGlnSerPheThrSerLeuValAsnSe 788

Qy       1154 GAAGATGGTAGTTCCT-----------------CCCAAACAGCAAATAAACCTTCAAGG 1195
                          :::                                :::|||:::|||
Db        788 rTyrGlnSerIleSerAsnLeuTyrHisGluLeuArgAspAspHisValAsnMetGln-- 807

Qy       1196 TTCCAAGCAAGAGACGCTG---------------------------------ACATT 1219
               |||:::       ||||||                                     ||
Db        808 -SerGlnAsnAsnThrLeuLeuGluSerGluSerLysLeuLysThrAspCysGluAsnLe 827

Qy       1220 GAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAA 1279
               |           |||    ||||||:::         |||:::   :::|||          :::
Db        827 uThrGlnGlnAsnMetThrLeuIleAspAsnValGln--LysLeuMetHisLysHisVal 846

Qy       1280 GGAGGAGAGAGTCGCACTGGA---ATGAATGATTTAAATGCTAAACTACCTGGAAATATT 1336
                       ||||||:::          :::     ::::::|||    ||||||        ::::::
Db        847 AsnGlnGluSerLysValSerGluLeuLysGluValAsnGlyLysLeuSerLeuAspLeu 866

Qy       1337 ------------TCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAG 1384
                          ||||||||||||||        ::::::    |||
Db        867 LysAsnLeuArgSerSerLeuAsnValAlaIleSerAspAsnAspGln------------ 882

Qy       1385 GATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAA 1444
                           |||          |||      |||   |||         |||          ::::::
Db        883 ---------IleLeuThrGlnLeuAlaGluLeuSerLysAsnTyrAspSerLeuGluGln 899

Qy       1445 CAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCAC 1504
                    |||:::         :::    |||
Db        900 ------GluSerAlaGlnLeuAsnSerGlyLeuLys--------------------Ser 910

Qy       1505 CTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGG 1564
              |||      :::|||:::   |||   |||      ||||||:::
Db        911 LeuGluAlaGluLysGlnLeuLeuHisThrGluAsnGluGluLeuHis------------ 926

Qy       1565 TGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTA 1624
                    |||||||||    |||                     |||   :::::::::
Db        927 ------IleArgLeuAspLysLeu------------------ThrGlyLysLeuLysIle 938

Qy       1625 GATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGG 1684
              :::       ::::::   ::::::     |||                 :::::::::
Db        939 GluGluSerLysSerSerAspLeuGlyLysLysLeuThrAlaArgGlnGluGluIleSer 958

Qy       1685 GATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAG 1744
              :::   |||          |||       |||   |||     :::         |||:::    :::
Db        959 AsnLeuLysGluGluAsnMetSerGlnSer---GlnAlaIleThrSerValLysSerLys 977

Qy       1745 AATGGTGAGCGA---ACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAAC 1801
                    |||      :::        |||:::    |||                 |||    |||
Db        978 LeuAspGluThrLeuSerLysSerSerLysLeuGluAlaAspIleGluHisLeuLysAsn 997

Qy       1802 AAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCT 1861
              |||    :::       :::         ::::::|||:::|||       |||
Db        998 LysValSerGluValGluValGluArgAsnAlaLeuLeuAlaSerAsn------------ 1013

Qy       1862 AAGAAATGGGACTTTACAGCTTTGATCGAT------------------------------ 1891
                             |||:::|||
Db       1014 ---------------GluArgLeuMetAspAspLeuLysAsnAsnGlyGluAsnIleAla 1028

Qy       1892 TTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCT 1951
                          :::|||:::        |||::::::
Db       1029 SerLeuGlnThrGluIleGluLysLysArgAlaGluAsn--------------------- 1041

Qy       1952 GATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTG 2011
              |||    ::::::        |||    :::    :::          :::    :::    |||:::
Db       1042 AspAspLeuGlnSerLysLeuSerValValSerSerGluTyrGluAsnLeuLeuLeuIle 1061

Qy       2012 AAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCT 2071
```

```
                 :::  |||   |||:::                   |||:::      :::
Db       1062 SerSerGlnThrAsnLysSerLeuGluAspLysThrAsnGlnLeuLysTyrIleGluLys 1081

Qy       2072 AATGCT------TTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTAT 2125
              |||                          :::||||||   :::::::::
Db       1082 AsnValGlnLysLeuLeuAspGluLysAspGlnArgAsnValGluLeuGluGluLeuThr 1101

Qy       2126 TCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCG 2185
              |||   |||                          |||:::         :::
Db       1102 SerLysTyr-----------------------------GlyLysLeuGlyGluGluAsn 1111

Qy       2186 GAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGG 2245
                 ::::::|||                     |||:::   ::::::|||:::
Db       1112 AlaGlnIleLysAspGluLeuLeuAlaLeuArgLysLysSerLysLysGlnHisAspLeu 1131

Qy       2246 TTGGTGACATTTACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAA 2305
                       |||            |||            |||:::
Db       1132 CysAlaAsnPheValAspAspLeuLysGluLysSerAspAlaLeuGluGlnLeuThrAsn 1151

Qy       2306 CCCTTGACACGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATG 2365
                       |||  |||:::   |||   :::         |||   :::   :::
Db       1152 GluLysAsnGluLeuIleValSerLeuGluGlnSer---AsnSerAsnAsnGluAlaLeu 1170

Qy       2366 CTACAG------GTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
              ::::::          |||   |||||||||
Db       1171 ValGluGluArgSerAspLeuAlaAsnArg----------------------------- 1180

Qy       2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
                 :::   :::      :::   |||   :::   :::   :::||||||   :::
Db       1181 ---LeuSerAspMetLysLysSerLeuSerAspSerAspAsnValIleSerValIleArg 1199

Qy       2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
              ::::::::::::      |||:::   |||                ::::::   |||
Db       1200 SerAspLeuValArgValAsnAspGluLeuAspThrLeuLysLysAspLysAspSerLeu 1219

Qy       2540 TACACAGGCTATGCTGAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGG 2599
                 |||   |||:::|||         :::                        :::|||
Db       1220 SerThrGlnTyrSerGluValCysGln--------------------------AspArg 1230

Qy       2600 GACGAC--------------------TGCGAGCGGCGC-------------------- 2617
              ||||||                    ||||||
Db       1231 AspAspLeuLeuAspSerLeuLysGlyCysGluGluSerPheAsnLysTyrAlaValSer 1250

Qy       2618 ------------TGCACT------------------------------------------ 2623
                          ||||||
Db       1251 LeuArgGluLeuCysThrLysSerGluIleAspValProValSerGluIleLeuAspAsp 1270

Qy       2624 ------------------------------GAGATCGTTGCCATCGATGCTCTTCACTTC 2653
                                            |||:::      :::      |||
Db       1271 AsnPheValPheAsnAlaGlyAsnPheSerGluLeuSerArgLeuThrValLeuSerLeu 1290

Qy       2654 AGACGCTACCTCGATCAGTTT 2674
                 |||||||||   |||
Db       1291 GluAsnTyrLeuAspAlaPhe 1297
```

RESULT 10

```
BMS1_YEAST
ID   BMS1_YEAST     STANDARD;      PRT;  1183 AA.
AC   Q08965;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ribosome biogenesis protein BMS1.
GN   BMS1 OR YPL217C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=21449426; PubMed=11565748;
RA   Wegierski T., Billy E., Nasr F., Filipowicz W.;
RT   "Bms1p, a G-domain-containing protein, associates with Rcl1p and is
RT   required for 18S rRNA biogenesis in yeast.";
RL   RNA 7:1254-1267(2001).
CC   -!- FUNCTION: May act as a molecular switch during maturation of the
CC       40S ribosomal subunit in the nucleolus. The depletion of BMS1
CC       interferes with processing of the 35S pre-rRNA at sites A0, A1,
CC       and A2, and the formation of 40S subunits.
CC   -!- SUBUNIT: Associates with RCL1.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: Belongs to the BMS1 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z73573; CAA97932.1; -.
DR   PIR; S65236; S65236.
DR   SGD; S0006138; BMS1.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005730; C:nucleolus; IDA.
DR   GO; GO:0006365; P:35S primary transcript processing; IDA.
DR   GO; GO:0042255; P:ribosome assembly; IDA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR007034; DUF663.
DR   Pfam; PF04950; DUF663; 1.
DR   SMART; SM00382; AAA; 1.
KW   Ribosome biogenesis; Nuclear protein; ATP-binding.
FT   NP_BIND      76     83       ATP (POTENTIAL).
SQ   SEQUENCE   1183 AA;  135570 MW;  9A337F1EE0B0F21D CRC64;
```

Alignment Scores:

Pred. No.:	0.00603	Length:	1183
Score:	159.50	Matches:	156
Percent Similarity:	31.89%	Conservative:	116
Best Local Similarity:	18.29%	Mismatches:	274
Query Match:	2.21%	Indels:	307
DB:	1	Gaps:	36

US-09-302-812-3 (1-4069) x BMS1_YEAST (1-1183)

```
Qy        263 AGCAGGCAGAGGCGCGTCCTCGACCCCAAGGAC--------------------------- 295
                 :::|||:::|||:::      ||||||   ||||||
Db        341 ThrArgArgArgLysArgLeuAspAspLysAspLysLeuIleTyrAlaProMetSerAsp 360

Qy        296 -----------------------------GCTCACGTGCAGTTCAGGGTCCCACCGTCC 325
                                             ::::::
Db        361 ValGlyGlyValLeuMetAspLysAspAlaValTyrIleAspIleGlyLysLysAsnGlu 380
```

```
Qy    326 TCGCCAGCCTGCGTCCCAGGGCAG---GCGGGACAGCACAGAGGCAGCGCCACCTCGCTT 382
Db    381 GluProSerPheValProGlyGlnGluArgGlyGluGlyGluLysLeuMetThrGlyLeu 400

Qy    383 GTTTTCAAACAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATC------------ 430
Db    401 GlnSerValGluGlnSerIleAlaGluLysPheAspGlyValGlyLeuGlnLeuPheSer 420

Qy    431 ------------------AAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAAACAAC 472
Db    421 AsnGlyThrGluLeuHisGluValAlaAspHisGluGlyMetAsp---ValGluSerGly 439

Qy    473 AATACAAGAATAGAA------------------TCCATGATGAGTTCTGTACAAAAAGAT 514
Db    440 GluGluSerIleGluAspAspGluGlyLysSerLysGlyArgThrSerLeuArgLysPro 459

Qy    515 AACTTTTACCAACATAATGTA---GAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGAT 571
Db    460 ArgIleTyrGlyLysProValGlnGluGluAspAlaAspIleAspAsnLeuProSerAsp 479

Qy    572 AAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGT 631
Db    480 GluGlu------------------ProTyrThrAsnAsp--------------------- 486

Qy    632 AAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTA 691
Db    487 ---------------------------AspAspValGlnAspSerGluProArgMetVal 497

Qy    692 ACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGAT 751
Db    498 GluIle---------AspPheAsnAsnThrGlyGluGlnGlyAlaGluLysLeuAlaLeu 514

Qy    752 CACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG 811
Db    515 GluThrAspSerGluPheGluGluSerGluAspGluPheSerTrpGluArgThrAla--- 533

Qy    812 CTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCC--------------------- 850
Db    534 ---AlaAsnLysLeuLysLysThrGluSerLysLysArgThrTrpAsnIleGlyLysLeu 552

Qy    851 ------AGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGAC 904
Db    553 IleTyrMetAspAsnIleSerProGluGluCysIleArgArgTrpArg---GlyGluAsp 571

Qy    905 TGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTT 964
Db    572 AspAspSer------LysAspGluSerAspIle--------------------------- 580

Qy    965 GGCTCTGAGGATGTT---------------------------GGTACTGGGTCAAAAAAT 997
Db    581 ---GluGluAspValAspAspAspPhePheArgLysLysAspGlyThrValThrLysGlu 599

Qy    998 GACAACAAA------TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAG 1051
Db    600 GlyAsnLysAspHisAlaValAspLeuGluLysPheValProTyrPheAspThrPheGlu 619

Qy   1052 AAGGAAAGTGAACCCGAATCACCGATGGATGTG--------------------------- 1084
Db    620 LysLeuAlaLysLysTrpLysSerValAspAlaIleLysGluArgPheLeuGlyAlaGly 639

Qy   1085 ---------GATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGT 1135
Db    640 IleLeuGlyAsnAspAsnLysThrLysSerAspSerAsnGluGlyGlyGluGluLeuTyr 659

Qy   1136 CCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
Db    660 GlyAspPheGluAspLeuGluAspGlyAsnProSerGluGlnAlaGluAspAsnSerAsp 679

Qy   1196 TTCCAAGCAAGAGACGCTGAC--------------------------------------- 1216
Db    680 LysGluSerGluAspGluAspGluAsnGluAspThrAsnGlyAspAspAspAsnSerPhe 699

Qy   1217 ---------------------------------------ATTGAATTTAGGAAACGGTAC 1237
Db    700 ThrAsnPheAspAlaGluGluLysLysAspLeuThrMetGluGlnGluArgGluMetAsn 719

Qy   1238 TCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACT 1297
Db    720 AlaAlaLysLysGluLysLeuArgAlaGlnPheGluIleGluGluGlyGluAsnPheLys 739

Qy   1298 GGAATGAATGATTTAAAT-----------------------------GCTAAACTACCT 1327
Db    740 GluAspAspGluAsnAsnGluTyrAspThrTrpTyrGluLeuGlnLysAlaLysIleSer 759

Qy   1328 GGAAATATTTCTAGCCTGAATGTAGAATGC---------------------AGAAATTCT 1366
Db    760 LysGlnLeuGluIleAsnAsnIleGluTyrGlnGluMetThrProGluGlnArgGlnArg 779

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCC------ 1420
Db    780 IleGluGlyPheLysAlaGlySerTyrValArgIleValPheGluLysValProMetGlu 799

Qy   1420 ------------------------------------------------------------ 1420
Db    800 PheValLysAsnPheAsnProLysPheProIleValMetGlyGlyLeuLeuProThrGlu 819

Qy   1421 ------------------AAAGCAGAGGACAGAAGAAAAGAACAGTGGGAA--------- 1453
Db    820 IleLysPheGlyIleValLysAlaArgLeuArgArg---HisArgTrpHisLysLysIle 838

Qy   1453 ------------------------------------------------------------ 1453
Db    839 LeuLysThrAsnAspProLeuValLeuSerLeuGlyTrpArgArgPheGlnThrLeuPro 858

Qy   1454 ---------ACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCAC 1504
Db    859 IleTyrThrThrThrAspSerArgThrArgThrArgMetLeuLysTyrThrProGluHis 878

Qy   1505 CTTTCTCCAGATAAGAAGTGGCTTGGA---------ACTCCCATTGAGGAGATGAGAAGA 1555
Db    879 ThrTyrCysAsnAlaAlaPheTyrGlyProLeuCysSerPro---------------Asn 893

Qy   1556 ATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT------------ 1603
Db    894 ThrProPheCysGlyValGlnIleValAlaAsnSerAspThrGlyAsnGlyPheArgIle 913

Qy   1604 ------------------CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAA 1645
Db    914 AlaAlaThrGlyIleValGluGluIleAspValAsnIleGluIleValLysLysLeuLys 933

Qy   1646 GTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAA--- 1702
Db    934 Leu---ValGlyPhePro------TyrLysIlePheLysAsnThrAlaPheIleLysAsp 950

Qy   1703 ATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCG 1762
Db    951 MetPheSerSerAlaMetGluValAlaArgPheGluGlyAlaGlnIleLysThrValSer 970

Qy   1763 GGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAAC 1822
Db    971 GlyIleArgGlyGlu---IleLysArgAlaLeu------------SerLysProGluGly 985

Qy   1823 ---------------TTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAA 1867
Db    986 HisTyrArgAlaAlaPheGluAspLysIleLeuMetSerAspIleValIleLeuArgSer 1005

Qy   1868 TGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAA 1927
Db   1006 TrpTyrProValArgValLysLysPheTyrAsnProValThrSer--------------- 1020

Qy   1928 CATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATT 1987
Db   1020 ------------------------------------------------------------ 1020

Qy   1988 TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAG 2047
```

```
Db     1021 ---------------LeuLeuLeuLysGluLysThrGluTrpLysGlyLeuArgLeuThr 1035

Qy     2048 GAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT 2107
               ||||||                                             |||
Db     1036 GlyGlnIle---------------------------------------------ArgAla 1040

Qy     2108 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTG--------- 2158
               |||       |||       |||    ||||||        :::::::::::
Db     1041 AlaMetAsnLeuGluThrProSerAsnProAspSerAlaTyrHisLysIleGluArgVal 1060

Qy     2159 ---------TTTGAGGGACGTTCA------------------------------------ 2173
                     |||   |||
Db     1061 GluArgHisPheAsnGlyLeuLysValProLysAlaValGlnLysGluLeuProPheLys 1080

Qy     2174 ------------TCAAGGAAACCGGAGAAACTTAAAACG 2200
                        ||||||:::|||   ||||||
Db     1081 SerGlnIleHisGlnMetLysProGlnLysLysLysThr 1093

RESULT 11
SR40_YEAST
ID   SR40_YEAST     STANDARD;      PRT;   406 AA.
AC   P32583;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Suppressor protein SRP40.
GN   SRP40 OR YKR092C OR YKR412A.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 28383 / FL100;
RA   Lalo D., Carles C., Sentenac A., Thuriaux P.;
RL   Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: Not known; weak suppressor of a mutant of the
CC       subunit AC40 of DNA dependant RNA polymerase I and III.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L11275; AAA35091.1; -.
DR   EMBL; X73541; CAA51946.1; -.
DR   EMBL; Z28317; CAA82171.1; -.
DR   PIR; S38170; S38170.
DR   GermOnline; 140071; -.
DR   SGD; S0001800; SRP40.
DR   GO; GO:0005730; C:nucleolus; IDA.
DR   InterPro; IPR007718; SRP40_C.
DR   Pfam; PF05022; SRP40_C; 1.
FT   DOMAIN       25    314       ASP/SER-RICH.
FT   CONFLICT    400    400       G -> N (IN REF. 1).
SQ   SEQUENCE   406 AA;  41015 MW;  8EA007695AF4BA1D CRC64;

Alignment Scores:
Pred. No.:             0.00491        Length:         406
Score:                 159.00         Matches:        61
Percent Similarity:    40.99%         Conservative:   80

Best Local Similarity:  17.73%       Mismatches:     179
Query Match:            2.21%        Indels:         24
DB:                     1            Gaps:           4

US-09-302-812-3 (1-4069) x SR40_YEAST (1-406)

Qy      281 CTCGACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTC 340
               |||      |||:::      ::::::   :::            ||||||   :::
Db       14 LeuSerValLysGluLysGluIleGluGluLysSerSerSerSerSerSerSerSerSer 33

Qy      341 CCAGGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACT 400
                     :::                     |||::::::|||         :::   :::
Db       34 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerGlyGluSerSer 53

Qy      401 ATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGT 460
               :::|||         :::            :::        |||:::|||   ||||||
Db       54 SerSerSerSerSerSerSerSerSerSerSerSerSerAspSerSerAspSerSerAspSer 73

Qy      461 AAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTT 520
            :::   ::::::::::::                |||      ||||||        |||:::
Db       74 GluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerAspSerGlu 93

Qy      521 TACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTC 580
                        :::         :::  |||        |||          |||
Db       94 SerSerSerGluSerAspSerSerSerSerGlySerSerSerSerSerSerSerSer 113

Qy      581 ACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAG---TGG 637
               |||   |||::::::         ::::::   :::|||            :::
Db      114 AspGluSerSerSerGluSerGluSerGluAspGluThrLysLysArgAlaArgGluSer 133

Qy      638 CAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAA------------ 685
               ||||||             :::      ::::::||||||:::
Db      134 AspAsnGluAspAlaLysGluThrLysLysAlaLysThrGluProGluSerSerSerSer 153

Qy      686 ------------------------ACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCT 721
                                    :::   :::         |||       |||   :::
Db      154 SerGluSerSerSerSerGlySerSerSerSerSerGluSerGluSerGlySerGluSer 173

Qy      722 AACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGA 781
            :::   |||   |||        :::         ||||||::::::|||:::
Db      174 AspSerAspSerSerSerSerSerSerSerSerSerAspSerGluSerAspSerGluSer 193

Qy      782 GACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGAT 841
            |||:::|||                               :::::::::   :::::::::
Db      194 AspSerGlnSer---------------------SerSerSerSerSerSerAspSer 206

Qy      842 GAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAA 901
                        ::::::   |||                  |||   |||
Db      207 SerSerAspSerAspSerSerSerSerAspSerSerAspSerAspSerSerSerSer 226

Qy      902 GACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGAT 961
                  :::|||            :::   |||            |||
Db      227 SerSerSerSerSerAspSerAspSerAspSerAspSerSerSerAspSerAspSer 246

Qy      962 GTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAA 1021
               ||||||   |||       :::   |||   :::   :::                  |||
Db      247 SerGlySerSerAspSerSerSerSerAspSerSerSer------------AspGlu 262

Qy     1022 AGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGAT 1081
            |||         :::|||         :::   :::|||:::       |||       :::
Db      263 SerThrSerSerAspSerSerAspSerAspSerAspSerAspSerGlySerSerSerGlu 282

Qy     1082 GTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGT 1141
            ::::::      :::   :::      :::|||:::|||:::|||      :::
Db      283 LeuGluThrLysGluAlaThrAlaAspGluSerLysAlaGluGluThrProAlaSerSer 302

Qy     1142 TTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAA 1201
               :::                  |||||||||   :::||||||||||   :::
Db      303 AsnGluSerThrProSerAlaSerSerSerSerSerAlaAsnLysLeuAsnIleProAla 322

Qy     1202 GCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGA 1261
                  |||      |||   ::::::::::::   ::::::                :::
Db      323 GlyThrAspGluIleLysGluGlyGlnArgLysHisPheSerArgValAspArgSerLys 342
```

```
Qy       1262 TTACATTTCCAA 1273
              ::::::|||:::
Db        343 IleAsnPheGlu 346

RESULT 12
YDG9_SCHPO
ID   YDG9_SCHPO     STANDARD;      PRT;  1031 AA.
AC   Q10496;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein C26F1.09 in chromosome I.
GN   SPAC26F1.09.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: Contains 1 Rab-GAP TBC domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z73100; CAA97366.1; -.
DR   PIR; T38411; T38411.
DR   GeneDB_SPombe; SPAC26F1.09; -.
DR   InterPro; IPR000195; RabGAP_TBC.
DR   Pfam; PF00566; TBC; 1.
DR   SMART; SM00164; TBC; 1.
DR   PROSITE; PS50086; TBC_RABGAP; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     798    818       POTENTIAL.
FT   DOMAIN       610    806       RAB-GAP TBC.
SQ   SEQUENCE   1031 AA;  116289 MW;  F2579A4C421D79EC CRC64;

Alignment Scores:
Pred. No.:             0.00628        Length:        1031
```

```
Score:                     159.00        Matches:        185
Percent Similarity:        30.89%        Conservative:   119
Best Local Similarity:     18.80%        Mismatches:     351
Query Match:               2.21%         Indels:         330
DB:                        1             Gaps:           39

US-09-302-812-3 (1-4069) x YDG9_SCHPO (1-1031)

Qy        392 CAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAG---ACAGCGGAATCAGAA 448
              ::::::   :::    |||      :::    :::    :::|||          |||   |||
Db          8 GluArgGluValGlnSerSerTyrGluLysGluAsnValLysIleIleArgGluGluGlu 27

Qy        449 AGTTTGGATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAA 508
              :::    |||    :::       ::::::       :::|||                :::
Db         28 AlaLysAspGlnGluSerThrAspAspIleAlaValGluAspGlyThrGlyThrSerPro 47

Qy        509 AAAGATAACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTT 568
                    ||||||:::                              ||||||     |||::::::
Db         48 AspLeuAsnPhePheSerThrGln---------------AsnValMetGlnMetAsnPhe 62

Qy        569 GATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATG 628
              :::          :::|||
Db         63 GluAspGluTyrSerGlu------------------------------------------ 68

Qy        629 TGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACA 688
                    :::     ||||||         |||
Db         69 ------PheSerAsnGluAspAspGluAlaGlu--------------------------- 77

Qy        689 GTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGAT 748
                                     |||:::          |||          |||
Db         78 ---------------------IleAspAsnSerPheAlaAspSerIleProAsnGluPro 90

Qy        749 GATCACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTA 808
              :::          |||          |||    :::||||||:::                :::
Db         91 GluIleProAspMetGlnAspGluTyrSerArgAspSerHisSerGlnGlnSerValGlu 110

Qy        809 AAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCAC 868
              :::          |||          |||    ||||||                |||
Db        111 GluGlnAsnAsnThrThrAsnThrAspGluAspAlaSerValAsnGluPheSerValAla 130

Qy        869 CAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAG 928
                    |||                   |||:::|||       :::             ::::::
Db        131 AlaAspIleSerAspValAsnThrLeuGlyLysAspAsnSerGluSerThrGluGluPro 150

Qy        929 ATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT------ 982
              ::::::    |||    ::::::             :::    :::|||:::||||||||||||
Db        151 ValAsnGluValAsnGluThr------AlaThrLeuGlyAsnGluAspValGlyGluArg 168

Qy        983 ---------------ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGC 1027
                                |||       |||:::    :::             |||
Db        169 SerGlyPheProSerGluGlyLeuAspAsnGluProGluSerGlnArgAspLeuAspGlu 188

Qy       1028 CTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCCGAATCACCG------------ 1075
                 ||||||       |||    :::    :::    |||    :::|||
Db        189 ThrGlyAsnLeuAlaProGluAspLeuLysAspGluValLysSerValHisGluPheAsn 208

Qy       1075 ------------------------------------------------------------ 1075

Db        209 GluProAsnAspLeuArgGlnGlnGluGluSerTyrSerAspAspAspAspThrAsnVal 228

Qy       1076 ---------ATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
                          ||||||:::       :::|||       |||    |||    |||||||||
Db        229 AsnGluPheGluAspValAsnGluIleGluAsnGluHisGlnLeuSerValAlaAspGlu 248

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATA-- 1184
              ::: |||          |||:::       ||||||:::          |||:::
Db        249 Asp-GlnThrSerArgLeuValLysGlyLysMetIlePheValGlyLysGluAspPheGl 268

Qy       1185 ----------------------------------------------------------AA 1186

Db        268 yGluGluAlaAspIleSerAsnSerValPheIleGluGlnAsnGlyProAsnSerAspTh 288

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
```

```
Qy        397 KRDSKITDHF--------MRISKSEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGT 448
                 | : :|            || :|     :: |   : :||  |   | | :    |
Db        616 SEGSMVGPNFNSFDQQENSYFSKGASKRVPSPNIQIQ--SQNLPKGIA--LEPRR----T 667

Qy        449 PIEEMRKMP--RCGIHLPSLRPSASHTVTVRVDLLRAGEVPKPFPTHY---KDLWDNKHV 503
              | :   | |  : | | |:                        ::| |   |:   :: |: : :
Db        668 PFQSETKKPELKHGTHQPAY---------------PKKIPSPTRKHFPAERNTWNRQKI 711

Qy        504 KMPCSEQNLYPVEDENGERTA-GSRWELIQTALLNKFTRP-QNLKDAILKYNVAYSKKWD 561
                  |   |   | :|||   : |||  :       :::| |  | |  :|      |  ||
Db        712 LPPLKED--YGRQDENLRHPSYGSRGNI----FYHEYTNPYHNEKSQYIK-----SNPWD 760
```

RESULT 11
DSPP_MOUSE

```
ID   DSPP_MOUSE     STANDARD;      PRT;   934 AA.
AC   P97399; O70567;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP-
DE   3) [Contains: Dentin phosphoprotein (Dentin phosphophoryn) (DPP);
DE   Dentin sialoprotein (DSP)].
GN   DSPP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Molar;
RX   MEDLINE=97150835; PubMed=8995371;
RA   MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;
RT   "Dentin phosphoprotein and dentin sialoprotein are cleavage products
RT   expressed from a single transcript coded by a gene on human chromosome
RT   4. Dentin phosphoprotein DNA sequence determination.";
RL   J. Biol. Chem. 272:835-842(1997).
RN   [2]
RP   SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RX   MEDLINE=98211969; PubMed=9545272;
RA   Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,
RA   D'Souza R.N., Kozak C.A., MacDougall M.;
RT   "Genomic organization, chromosomal mapping, and promoter analysis of
RT   the mouse dentin sialophosphoprotein (Dspp) gene, which codes for both
RT   dentin sialoprotein and dentin phosphoprotein.";
RL   J. Biol. Chem. 273:9457-9464(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RA   Sfeir C., Butler S., Lin E., George A., Veis A.;
RT   "From mouse to zebrafish-dentin matrix proteins genomic
RT   characterization.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=98055479; PubMed=9395101;
RA   Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
RA   Hotton D., Papagerakis P., Berdal A., Butler W.T.;
RT   "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
RT   expression in odontoblasts and transient expression in
RT   pre-ameloblasts.";
RL   Eur. J. Oral Sci. 105:405-413(1997).
RN   [5]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=21096982; PubMed=11175790;
RA   Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,
RA   Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,
RA   Zhao G., Kong X.;
RT   "Dentinogenesis imperfecta 1 with or without progressive hearing loss
RT   is associated with distinct mutations in DSPP.";
RL   Nat. Genet. 27:201-204(2001).
CC   -!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
CC       may bind high amount of calcium and facilitate initial
CC       mineralization of dentin matrix collagen as well as regulate the
CC       size and shape of the crystals.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and
CC       transiently in pre-ameloblasts. Found in the inner ear.
CC   -!- PTM: DSP is glycosylated.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U67916; AAC12787.1; -.
DR   EMBL; AJ002141; CAA05208.1; ALT_SEQ.
DR   EMBL; AF135799; AAD42781.1; ALT_SEQ.
DR   MGD; MGI:109172; Dspp.
KW   Biomineralization; Extracellular matrix; Signal; Glycoprotein;
KW   Phosphorylation; Sialic acid.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    934       DENTIN SIALOPHOSPHOPROTEIN.
FT   CHAIN        18    451       DENTIN SIALOPROTEIN.
FT   CHAIN       452    934       DENTIN PHOSPHOPROTEIN.
FT   DOMAIN      419    934       SER/ASP-RICH.
FT   SITE        479    481       CELL ATTACHMENT SITE (POTENTIAL).
FT   MOD_RES     227    227       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     254    254       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     279    279       PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT   MOD_RES     293    293       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   MOD_RES     299    299       PHOSPHORYLATION (BY CK1) (BY SIMILARITY).
FT   MOD_RES     314    314       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     336    336       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   MOD_RES     349    349       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    313    313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    373    373       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   934 AA;  93901 MW;  A618789D8A57249A CRC64;

  Query Match             3.0%;  Score 154;  DB 1;  Length 934;
  Best Local Similarity   20.5%;  Pred. No. 0.16;
  Matches   80;  Conservative   60;  Mismatches  154;  Indels   96;  Gaps   14;

Qy         34 RQRRVLDPKDA-PVQFRVPPSSPACVSGRA--------GPHRGNATSFVFKQKTITTWMD 84
              : :  | |||    :   :    | || ||::        ||::||        |:|  |
Db        330 QDKEKLSPKDTRDAEGGIISQSEACPSGKSQDQGIETEGPNKGN--------KSIIT--K 379

Qy         85 TKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKSPTEKSSQYLN 144
                |   :  :|   :      :|    |  | ::  |                :   |||: :  |
Db        380 ESGKLSGSKDSNGHQGVELDKRNSPKQGESDKP--------------QGTAEKSAAHSN 424

Qy        145 QQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHSDTDHEEDRDN 204
                                                | ::  ||  | ::| |      :| |  ::
Db        425 ---------------------------------LGHSRIGSSSNSDGHD--SYEFDDES 448

Qy        205 QQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQEEVDVLPESPLSDVGAED 264
                |   |     :: : |  :: :| : :     |  | :      |   |     :|   |   | :|
Db        449 MQGDDP---KSSDESNGSDESDTNSESANESGSRGDASYTSDESSDDDNDSD-SHAGEDD 504

Qy        265 IGTGPKNDNKLTGQESSLGDSPPFEKESEPESPMDVDNSRNSCQDSEADEETSPVFDEQD 324
                    :|:   |   | ||    : |||  :   |  :   ||   |||:   ::|     |:
Db        505 ----SSDDSSGDGDSDSNGDG---DSESEDKDESDSSDHDNS-SDSESKSDSSDSSDDSS 556

Qy        325 DRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSSS 384
              | |  : :  ||  :  :|           :  |:             ::|:  :   ::  | :||
Db        557 DSSDSSDSSDSSDSSDSSDS-------SDSSD-------SSDSNSSSDSSDSSGSSDSSD 602
```

```
RT   "The complete cDNA sequence of bovine coagulation factor V.";
RL   J. Biol. Chem. 267:2971-2978(1992).
RN   [2]
RP   DISULFIDE BONDS.
RX   MEDLINE=95034740; PubMed=7947716;
RA   Xue J., Kalafatis M., Silveira J.R., Kung C., Mann K.G.;
RT   "Determination of the disulfide bridges in factor Va heavy chain.";
RL   Biochemistry 33:13109-13116(1994).
CC   -!- FUNCTION: Coagulation factor V is a cofactor that participates
CC       with factor Xa to activate prothrombin to thrombin.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC       chain, noncovalently bound. The interaction between the two chains
CC       is calcium-dependent.
CC   -!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
CC       AA repeats.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC       cofactor, factor Va (formation of a heavy chain at the N-
CC       terminus and a light chain at the C-terminus).
CC   -!- PTM: Sulfation is required for efficient thrombin cleavage and
CC       activation and for full procoagulant activity (By similarity).
CC   -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC   -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC   -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M81440; AAA30512.1; -.
DR   EMBL; M81441; AAA30513.1; -.
DR   PIR; A42580; KFBO5.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR008972; Cupredoxin.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR008979; Gal_bind_like.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS50022; FA58C_3; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
KW   Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW   Repeat.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   2211       COAGULATION FACTOR V.
FT   CHAIN        29    741       COAGULATION FACTOR HEAVY CHAIN (BY
FT                                SIMILARITY).
FT   PROPEP      742   1564       ACTIVATION PEPTIDE (CONNECTING REGION)
FT                                (BY SIMILARITY).
FT   CHAIN      1565   2211       COAGULATION FACTOR LIGHT CHAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       30    327       F5/8 TYPE A 1.
FT   DOMAIN       30    193       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      203    327       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      348    686       F5/8 TYPE A 2.
FT   DOMAIN      348    525       PLASTOCYANIN-LIKE 3.
FT   DOMAIN      535    686       PLASTOCYANIN-LIKE 4.
FT   DOMAIN      696   1564       B.
FT   DOMAIN     1124   1151       2 X 14 AA TANDEM REPEATS.
FT   REPEAT     1124   1137       1-1.
FT   REPEAT     1138   1151       1-2.
FT   DOMAIN     1188   1453       30 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                                [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT   REPEAT     1188   1196       2-1.
FT   REPEAT     1197   1205       2-2.
FT   REPEAT     1206   1214       2-3.
FT   REPEAT     1215   1223       2-4.
FT   REPEAT     1224   1232       2-5.
FT   REPEAT     1233   1241       2-6.
FT   REPEAT     1242   1250       2-7.
FT   REPEAT     1251   1259       2-8.
FT   REPEAT     1260   1268       2-9.
FT   REPEAT     1269   1277       2-10.
FT   REPEAT     1278   1286       2-11.
FT   REPEAT     1287   1295       2-12.
FT   REPEAT     1296   1304       2-13.
FT   REPEAT     1305   1313       2-14.
FT   REPEAT     1314   1322       2-15.
FT   REPEAT     1323   1331       2-16.
FT   REPEAT     1332   1340       2-17.
FT   REPEAT     1341   1349       2-18.
FT   REPEAT     1350   1358       2-19.
FT   REPEAT     1359   1367       2-20.
FT   REPEAT     1368   1376       2-21.
FT   REPEAT     1377   1385       2-22.
FT   REPEAT     1386   1394       2-23.
FT   REPEAT     1395   1403       2-24.
FT   REPEAT     1404   1412       2-25.
FT   REPEAT     1413   1421       2-26.
FT   REPEAT     1422   1430       2-27.
FT   REPEAT     1431   1439       2-28.
FT   REPEAT     1440   1444       2-29 (PARTIAL).
FT   REPEAT     1445   1453       2-30.
FT   DOMAIN     1569   1890       F5/8 TYPE A 3.
FT   DOMAIN     1569   1738       PLASTOCYANIN-LIKE 5.
FT   DOMAIN     1748   1890       PLASTOCYANIN-LIKE 6.
FT   DOMAIN     1894   2048       F5/8 TYPE C 1.
FT   DOMAIN     2053   2208       F5/8 TYPE C 2.
FT   SITE        741    742       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1034   1035       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1564   1565       CLEAVAGE (BY THROMBIN).
FT   DISULFID    167    193
FT   DISULFID    248    329
FT   DISULFID    499    525
FT   DISULFID    607    688
FT   DISULFID   1712   1738       PROBABLE.
FT   DISULFID   1894   2048       BY SIMILARITY.
FT   DISULFID   2053   2208       BY SIMILARITY.
FT   MOD_RES     697    697       SULFATION (POTENTIAL).
FT   MOD_RES     701    701       SULFATION (POTENTIAL).
FT   MOD_RES     730    730       SULFATION (POTENTIAL).
FT   MOD_RES    1513   1513       SULFATION (POTENTIAL).
FT   MOD_RES    1529   1529       SULFATION (POTENTIAL).
FT   MOD_RES    1537   1537       SULFATION (POTENTIAL).
FT   MOD_RES    1541   1541       SULFATION (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    587    587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    745    745       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    756    756       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    902    902       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    952    952       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1044   1044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1053   1053       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1062   1062       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1078   1078       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1094   1094       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1451   1451       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1490   1490       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1550   1550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1690   1690       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1839   1839       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1997   1997       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2196   2196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     587    592       NFTLPA -> T (IN VARIANT 2).
SQ   SEQUENCE   2211 AA;  248981 MW;  CBBF90B738667C45 CRC64;

  Query Match             3.0%;  Score 157.5;  DB 1;  Length 2211;
  Best Local Similarity   18.7%;  Pred. No. 0.32;
  Matches  135;  Conservative  102;  Mismatches  295;  Indels  189;  Gaps   28;

Qy       65 HRGNATSFVFKQKTITTWMDTKGP---------------------------------- 88
Db      641 HEDTLTLFPMQGESVTVTMDNVGTWMLTTMNSNPRSKKLRLRFRDAKCIRNDDDDSYEII 700

Qy       89 -----KTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKL------------ENVPQLNL 131
Db      701 YEPSGSTAMTTKKIHDSSEIEDENDA---DSDYQDELALILGLRSFRNSSLNQEKDELNL 757

Qy      132 DKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASE------------PPAGTPLPKQ 178
Db      758 TALALEKDSEFIPPSANRSLDSNSSSRSHVSRLIAKNFAESLKTLLHLEAPAAGSPLEHA 817

Qy      179 LSNANIGQSPHTDDHSDTDHEEDRDNQQF--LTPIKL-------ANTKPT------VGDG 223
Db      818 GLDKNSALNPPMAEHSSPYSEDPREDHPLSDVTGVSLLPFGTGFKNRKPAKHQRFQVGRG 877

Qy      224 QARSNCKCSGSR-QSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNK-LTGQESS 281
Db      878 QAAKH-KFSQTRFPAHKTRTRLSQDNSSSSRMGPWEDIPSDLLLLQQKDPYKILNGEWHL 936

Qy      282 LGDSPPFEKESEPESPMDVDNSRNSCQ------------DSEADEETSPVF-------- 320
Db      937 VSEKGSYEIIQDANENKTVNKLPNSPQNDSRTWGENIPFKNSHGKQSGHPTFLVTRRKPL 996

Qy      321 -DEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPS 379
Db      997 QDRQDRRNSRLKEGLPLIRTRRKKKEEKPAYHVPLSPRSFH-PLRGEVNASFSDRRHNHS 1055

Qy      380 -----GNSSSLNVECRSS--KQHGKRDSKITDHFMRISKSEDRRKEQCEVRHQRTERKIP 432
Db     1056 LLLHASNETSLSIDLNQTFPSMNLSLAASLPDH-----------DQTSPNDTTSQTSSP 1103

Qy      433 KYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLR-PSASHTVTV---RVDLLRAGEVPK 488
Db     1104 PDLYPTVSPEEHYQIFPIQD--SDPTHSTTAPSNRSPDPTHSTTAPSNRSPPTQPSQIPN 1161

Qy      489 PFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKDA 548
Db     1162 ---------YDLRNRAIPTDVSQIFP---------SLELEVWQTATSLDLSQP------ 1196

Qy      549 ILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPI-PLL 607
Db     1197 --------------SISPDLGQMALSPDPGQ---ESLSPDLGQTSLS-PDLSQESLSPDL 1238

Qy      608 KQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLF--EGRSSRKP 665
Db     1239 GQTA-LSPDPSQESLSPDLGQTALSPDP-------SQESLSPDLGQTALSPDPGQESLSP 1290

Qy      666 E 666
Db     1291 D 1291

RESULT 10
ENAM_MOUSE
ID   ENAM_MOUSE     STANDARD;      PRT;  1274 AA.
AC   O55196;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Enamelin precursor.
GN   ENAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Enamel epithelium;
RX   MEDLINE=20514993; PubMed=11062988;
RA   Hu C.-C., Simmer J.P., Bartlett J.D., Nanci A., Qian Q., Zhang C.,
RA   Ryu O.H., Xue J., Fukae M., Uchida T., McDougall M.;
RT   "Murine enamelin: cDNA and derived protein sequences.";
RL   Connect. Tissue Res. 39:47-61(1998).
CC   -!- FUNCTION: Involved in the mineralization and structural
CC       organization of enamel. Involved in the extension of enamel during
CC       the secretory stage of dental enamel formation.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC   -!- TISSUE SPECIFICITY: Expressed in developing teeths.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U82698; AAB94312.1; -.
DR   PIR; T37193; T37193.
DR   MGD; MGI:1333772; Enam.
DR   GO; GO:0005578; C:extracellular matrix; ISS.
DR   GO; GO:0030345; F:structural constituent of tooth enamel; ISS.
DR   GO; GO:0030282; P:bone mineralization; ISS.
DR   GO; GO:0042476; P:odontogenesis; ISS.
KW   Biomineralization; Extracellular matrix; Glycoprotein; Signal.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1274       ENAMELIN.
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    252    252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    300    300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1066   1066       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1274 AA;  140954 MW;  F9DBD1CC9D327143 CRC64;

  Query Match             3.0%;  Score 156;  DB 1;  Length 1274;
  Best Local Similarity   21.2%;  Pred. No. 0.19;
  Matches  127;  Conservative   74;  Mismatches  239;  Indels  160;  Gaps   32;

Qy       41 PKDAPVQFRVPPSSPACVSGRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESESKENNN 100
Db      242 PDDPPPEASTNSTVPDANATQSIPEGGNDTSPI-------------GNTGPGPNAGNNP 287

Qy      101 TRIDSMMSSVQKDNFYPHKVEK-----------------LENVPQLNL---------- 131
Db      288 T--------VQNGVFPPPKVNVSGQGVPKSQIPWRPSQPNIYENYPYPNYPSERQWQTTG 339

Qy      132 DKSPTEKSSQYLNQQQTASVCKWQN---EGKHAEQLLASEPPAGTPLPKQLSNANIGQSP 188
Db      340 TQGPRQNGPGYRNPQVERGP-QWNSFAWEGKQATR--PGNPTYGKP-PSPTSGVNYAGNP 395

Qy      189 HTDDHSDTDHEEDRDNQQFL--TPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQ 246
Db      396 -----VHFGRNLPGPNKPFVGANP---ASNKPFVGANPASNKPFVGANPASNKPFVGAN- 446

Qy      247 EEVDVLPESPLSDVGAED------IGTGPKNDNKLTG-----QESSLGDSPPFEKESEPE 295
Db      447 ------PASNKPFVGANPASNKPYVGANPASNKPFIGANPAANKPSIGTNPAANKPSIGT 500

Qy      296 SP-MDVDNSRNSCQDSEADEETSPVFDEQDDRSSQTANK--LSSCQAREAD------GDL 346
Db      501 NPAANKPFVRNNVGANKPFVGTNPSSNQPFLRSNQASNKPFMRSNQASNKPFVGTNVASV 560

Qy      347 RKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVECRSSKQHG---------- 396
Db      561 GPKQVTVSHNMKT--QNPKEKSLGQKERTVTPTKDASN---PWRSAKQYGINNPNYNLPR 615
```

```
Qy        385 LNVECRSSKQHGKRDSKITDHFMRISKSED 414
                 : | ||      || :     | | |
Db        603 SSDTCDSSDSSDSSDSSDSSDSSDSSDSSD 632

RESULT 12
PCLO_RAT
ID   PCLO_RAT       STANDARD;      PRT;  5085 AA.
AC   Q9JKS6; Q9JLT1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN   PCLO.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX   MEDLINE=20170257; PubMed=10707984;
RA   Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA   Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT   "Piccolo, a presynaptic zinc finger protein structurally related to
RT   bassoon.";
RL   Neuron 25:203-214(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP   VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP   ALA-4694.
RX   MEDLINE=21181819; PubMed=11285225;
RA   Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT   "An unusual C(2)-domain in the active-zone protein piccolo:
RT   implications for Ca(2+) regulation of neurotransmitter release.";
RL   EMBO J. 20:1605-1619(2001).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC       junctions.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9JKS6-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF138789; AAF07822.2; -.
DR   EMBL; AF227534; AAF63196.1; -.
DR   HSSP; P04410; 1A25.
DR   GO; GO:0045202; C:synaptic junction; IDA.
DR   GO; GO:0005509; F:calcium ion binding; IDA.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW   Repeat; Alternative splicing.
FT   DOMAIN      372    491       12 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     523    547       C4-TYPE (POTENTIAL).
FT   ZN_FING    1010   1033       C4-TYPE (POTENTIAL).
FT   DOMAIN     2351   2362       POLY-PRO.
FT   DOMAIN     4442   4536       PDZ.
FT   DOMAIN     4653   4752       C2 DOMAIN 1.
FT   DOMAIN     4968   5059       C2 DOMAIN 2.
FT   VARSPLIC   4876   4880       TKPTN -> SKRRK (in isoform 2).
FT                                /FTId=VSP_003930.
FT   VARSPLIC   4881   5085       Missing (in isoform 2).
FT                                /FTId=VSP_003931.
FT   MUTAGEN    4668   4668       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                BINDING ACTIVITY.
FT   MUTAGEN    4674   4674       D->A: COMPLETE LOSS OF CALCIUM-BINDING
FT                                AND CALCIUM-DEPENDENT PHOSPHOLIPID
FT                                BINDING ACTIVITY.
FT   MUTAGEN    4688   4688       V->S: SMALL INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4688   4689       VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4689   4689       M->S: INCREASED AFFINITY FOR CALCIUM.
FT   MUTAGEN    4690   4691       VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4692   4693       QN->AA: MODERATE INCREASE IN AFFINITY FOR
FT                                CALCIUM.
FT   MUTAGEN    4694   4694       A->S: NO EFFECT ON CALCIUM-BINDING
FT                                ACTIVITY.
SQ   SEQUENCE   5085 AA;  552702 MW;  5A1BB543201A7450 CRC64;

  Query Match             3.0%;  Score 153;  DB 1;  Length 5085;
  Best Local Similarity   20.2%;  Pred. No. 1.8;
  Matches  122;  Conservative   78;  Mismatches  225;  Indels  178;  Gaps   26;

Qy         75 KQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKS 134
              :||| |     :|       :| | ||:  ::     | |     :| :  |   |:|  | |
Db       1397 QQKTQTA-SETLDITISEEEIKESQEKKVSPKKDSEQG---FPSRKEHKEK-PELVDDLS 1451

Qy        135 PTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHS 194
              |    |  :                           |    :|: ::    :|| |   :|::
Db       1452 PRRASYDSVED--------------------SSESENSPVVRRKRRTSIGSS-SSDEYK 1489

Qy        195 DTDHEEDRDNQQFLTP--IKLANTKPT-----------------VGDGQARSNCKCSG- 233
                | :   : : |:     |::: :                   ||: |  |    |
Db       1490 QEDSQGSGEEEDFIRKQIIEMSADEDASGSEDEEFIRSQLKEISGVGESQKREEAKGKGK 1549

Qy        234 ------------SRQSVKDCTGCQ-------QEEVDVLPESPLSDVGAED------IGTG 268
                          |   |   |  | :            |  |  ||    :  :::      ||
Db       1550 GVAGKHRRLTRKSSTSFDDDAGRRHSWHDEDDETFDESPELKFRETKSQESEELVVAGGG 1609

Qy        269 PKNDNKLTGQESSLGD-----------------SPPFEKESEPESPMDVDNSR--NSCQD 309
                   |  |:: |                     |  |  ||  :||  |      :|
Db       1610 GLRRFKTIELNSTIADKYSSESSQKKTILYFDEEPELEMESLTDSPEDRSRGEGSSSLHA 1669

Qy        310 SEADEETSPV----FDEQDDRSSQTANKLSSCQAREADGDLRKRYLTKGSEVRLHFQFEG 365
              |     |||       ||  |  ||  :   |   |  | |:|       |: :  |   :
Db       1670 SSFTPGTSPTSVSSLDEDSD-SSPSHKKGESKQQRKA------RHRSHGPLL-------- 1714
```

```
Qy        366 ENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSK--ITDHFMRISKSEDRRKEQCEVR 423
                          |:  ||  | |  ::  |   |   :   | | |:  :|:: |:|
Db       1715 ------------PTIEDSSEEEELREEEELLKEQEKQRELEQQQRKSSSKKSKKDKDELR 1762

Qy        424 HQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPR--------CGIHLPSLRPSASHTVT 475
                || |:  ||  | ||  | :    :|  ||:|:          |  : ||:
Db       1763 AQR-RRERPKTPPSNLSPIED--ASPTEELRQAAEMEELHRSSCSEYSPSIESD------ 1813

Qy        476 VRVDLLRAGEVPKPFPTHYKDLWD-NKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTA 534
                          |: |     : : :  |   |:| :     | ::::  :   |      :| |
Db       1814 -----------PEGFEISPEKIIEVQKVYKLPAAVSLYSPTDEQSVMQKEG-----VQKA 1857

Qy        535 LLNKFTRPQNLKDAILKYNVAYSKKWDFTAL--VDFWDKVLEEAE-------AQHLYQSI 585
                         || |   |     |     :  |    :  |:| |:            ::|:|:
Db       1858 ----------LKSAEEMYEEMMQKPHKYKAFPAANERDEVFEKEPLYGGMLIEDYIYESL 1907

Qy        586 LPD 588
              : |
Db       1908 VED 1910
```

```
RESULT 13
SR40_YEAST
ID   SR40_YEAST     STANDARD;      PRT;   406 AA.
AC   P32583;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Suppressor protein SRP40.
GN   SRP40 OR YKR092C OR YKR412A.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 28383 / FL100;
RA   Lalo D., Carles C., Sentenac A., Thuriaux P.;
RL   Submitted (MAY-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: Not known; weak suppressor of a mutant of the
CC       subunit AC40 of DNA dependant RNA polymerase I and III.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L11275; AAA35091.1; -.
DR   EMBL; X73541; CAA51946.1; -.
DR   EMBL; Z28317; CAA82171.1; -.
DR   PIR; S38170; S38170.
DR   GermOnline; 140071; -.
DR   SGD; S0001800; SRP40.
DR   GO; GO:0005730; C:nucleolus; IDA.
DR   InterPro; IPR007718; SRP40_C.
DR   Pfam; PF05022; SRP40_C; 1.
FT   DOMAIN       25    314       ASP/SER-RICH.
FT   CONFLICT    400    400       G -> N (IN REF. 1).
SQ   SEQUENCE   406 AA;  41015 MW;  8EA007695AF4BA1D CRC64;

  Query Match             3.0%;  Score 152.5;  DB 1;  Length 406;
  Best Local Similarity   20.3%;  Pred. No. 0.066;
  Matches   75;  Conservative   58;  Mismatches  163;  Indels   73;  Gaps   12;

Qy         75 KQKTITTWMDTKGPKTAESESKENNNTRIDSMMSSVQKDNFYPHKVEKLENVPQLNLDKS 134
              |:| |    :  |   :: | |  ::::   |  ||   ::              :  : | |
Db         17 KEKEI----EEKSSSSSSSSSSSSSSSSSSSSSGESSSSSSSSSSSSSD-SSDSS 71

Qy        135 PTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLPKQLSNANIGQSPHTDDHS 194
               :| ||   :    ::|                   :|:  :  :     | ::    |  :|: |
Db         72 DSESSSSSSSSSSSS--------------SSSDSESSSESDSSSGSSSSSSSSSDESS 117

Qy        195 DTDHEEDR--------DNQQFLTPIKLANTKPTVGDGQARSNCKCSGSRQSVKDCTGCQQ 246
                   ||          ||:     | | |:|    :   :  |     |||  |  :   :|  :
Db        118 SESESEDETKKRARESDNED-AKETKKAKTEP---ESSSSESSSSGSSSSESESGSES 173

Qy        247 E-----------------EVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFE 289
              :                  | |   |  |   :        :|:         ||   |    :
Db        174 DSDSSSSSSSSSDSESDSESDSQSSSSSSSSDSSSDSDSSSSDSSSDSDSSSSSSSSSD 233

Qy        290 KESEPESPMDVDNS----RNSCQDSEADEETSPVFDEQDDRSSQTANKLSSCQAREADGD 345
               :|: :|  | |:|      :|  || :|| ||     |   |    |     :    |   : :||  |
Db        234 SDSDSDSSSDSDSSGSSDSSSSSDSSSDESTSS--DSSDSDSDSDSGSSSELETKEATAD 291

Qy        346 LRKRYLTKGSEVRLHFQFEGENNAGTSDLNAKPSGNSSSLNVECRSSK-QHGKRDSKITD 404
                |   |   |                  :   |   :|   | :::  ||:    : : :  |:|
Db        292 ESKAEETPAS-------------SNESTPSASSSSANKLNIPAGTDEIKEGQR-----K 333

Qy        405 HFMRISKSE 413
               || |: :|:
Db        334 HFSRVDRSK 342

RESULT 14
PCLO_CHICK
ID   PCLO_CHICK     STANDARD;      PRT;  5120 AA.
AC   Q9PU36;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragment).
GN   PCLO OR ACZ.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1 and profilin (By similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
                     :::   |||   |||:::   :::        :::|||      :::   |||
Db        288 rValSerGlyPheLysGluThrSerSerIleValAsnSerSerSerThrThrGluLysPr 308

Qy       1247 GGCGGTGAAGTTAGATTACAT----------------TTCCAATTTGAAGGAGGAGAG 1290
                 |||   ||||||                        |||      |||          ||
Db        308 oGlyValAlaLeuAspSerGlnAsnAspThrSerIlePheAsnGluGluGlnThrAsnSe 328

Qy       1291 TCGCACT---GGAATGAATGATTTAAATGCT---AAACTACCTGGAAATATTTCTAGC-- 1342
              |   |||            |||||||||         ||||||    |||:::    |||
Db        328 rLeuThrGluThrPheAsnAspLeuThrLeuAspHisLeuProGluAsnValGluSerGl 348

Qy       1343 ----------------CTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGA 1386
                                     |||           :::|||              ||
Db        348 uProValAlaGlyLysGluAsnGluThrAlaLysAsnGluSerGlyAlaSerAspAsnAs 368

Qy       1387 TTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACA 1446
              |   |||         |||:::       |||       :::||||||
Db        368 pHisLysAlaAsnValHisValPheValLeuLysSerSerGluAsp-------------- 383

Qy       1447 GTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCT 1506
                                         |||   ||||||              |||
Db        384 ----------AlaIleThrLeuAsnGluGluLysIleAlaThrGlnAspAspProLeuGl 400

Qy       1507 TTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTG 1566
                :::|||                      |||||||||
Db        400 uAlaPro------------------ThrProIle------------------------- 405

Qy       1567 TGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATT-------- 1618
                                          :::::::::      |||:::   :::
Db        406 -------------------------ValAlaSerSerSerThrIlePheLeuAsnSerAs 417

Qy       1619 ----CGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAA 1674
                    |||   |||   |||   |||            :::|||   |||         ||
Db        417 nGlnArgAsnAspGluLeuSerAlaSerGlySerGlnGluProHisPro---------Ly 434

Qy       1675 AGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGT 1734
               ||||          |||          :::|||              ||||||
Db        434 sAsp---GlyThrAsnSerThrSerSerLeuProLeuAspThrAsnAsnLeuSerAsnSe 453

Qy       1735 GGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGG---GAGCTCATTCAGACTGC 1791
               |||        :::            |||    |||            |||:::|||||||||
Db        453 rGluProProSerHisValLeuAspAlaSerSerGluThrIleGluValIleGlnThrIl 473

Qy       1792 ACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGT 1851
                      :::            |||:::    :::||||||    :::    |||    |||
Db        473 eLysLysLeuGlnAsnGlnValProGluThrIleLysAspGluValGlyLysLysAsnTh 493

Qy       1852 GGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGA 1911
                 |||:::|||            :::                    :::    |||
Db        493 rAlaPheSerProGlyThrSerLeuSerThr---------------AsnHisValLysTh 508

Qy       1912 AGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATTGCACT 1971
               ::::::   :::   |||
Db        508 rLysSerArgSerAlaHis------------------------------------------ 514

Qy       1972 CTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTC 2031
                                                                   |||:::||
Db        515 -------------------------------------------------------AsnAsnSe 517

Qy       2032 CATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATT 2091
              |   :::                       ::::::|||   |||
Db        517 rThrSerProPheSerThrAlaValSerSerTrpLeuAsn-------------------- 530

Qy       2092 TCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAA 2151
                 |||   |||                                  ||||||
Db        531 -ProLeuArg-----------------------------TyrPro-------------- 535

Qy       2152 TCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTA 2211
                                    |||    :::    |||          :::   :::      ||
Db        536 ------------------SerAspLysSerPro---------ArgValIleSerSerTy 546

Qy       2212 CTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAGAGTCT 2271
              |           |||          ||||||                             :::|||::
Db        546 rLeuGluSerValPheIleSerLysPro------------------------ArgSerIl 558

Qy       2272 TGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTA 2331
               :   |||       :::   |||     |||
Db        558 eGlyAspAlaGlnLysLeuGluIleLeuGlu---------------------------- 568

Qy       2332 CGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTT 2391
                                                |||||
Db        569 ------------------------------TyrLeuGln-------------------- 571

Qy       2392 TGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATC-------------- 2437
                                  :::   :::        |||      ::::::
Db        572 ----------------SerGlnSerSerThrValSerAsnGlnValPheThrLeuLeuSe 586

Qy       2438 -CGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAG----------- 2485
                  |||:::   ||||||        ::::::           |||     |||
Db        586 rAsnPheIleGlnAsnProLeuPheValLeuAspGluCysPheAspGluPheArgAsnLe 606

Qy       2486 -GTGCTGGATCACAATGAATGTCTAATTATCACAGGTACT---GAGCAGTACAGTGAATA 2541
                 :::|||   ||||||        :::   |||        :::      |||   ::
Db        606 uIleLeuMetHisAsnSerHisThrValHisThrValValTrpLysThrIleSerSerTr 626

Qy       2542 CACAGGCTATGCTGAGACATATCGTTGGTCC---------------CGGAGCCACGAAGA 2586
              :|||   |||           ::::::|||                         ||
Db        626 pThrSerTyrAspTyrGluMetGlnTyrSerSerLeuSerIleLysAsnCysAspSerAs 646

Qy       2587 TGGGAGTGAAAGGGACGACTGCGAGCGGCGC---TGCACTGAGATCGTTGCCATCGATGC 2643
              |   :::   |||   |||   :::|||                ||||||::::::
Db        646 pLysAlaIleArgLysAspLeuAspArgThrPheAlaProGluIleLeuSer-------- 663

Qy       2644 TCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAA 2703
                   ||||||               |||       |||               |||
Db        664 ----HisPhePheSerAsnArgGlnGlnLeuGluProThrAspAsnIleAlaGlu----- 680

Qy       2704 CAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGC 2763
                                              |||:::   ||||||      |||
Db        681 ----------------------------------SerThrAlaAsnLeuHisArgValLe 689

Qy       2764 CACAGGA------------------------------------------AACTGGGGCTG 2781
                                                                 :::|||
Db        689 uArgSerLeuAlaIleValLeuProGlnValGlyTyrThrGlnGlyMetSerTrpIleAl 709

Qy       2782 TGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGC 2841
               |||                          ||||||:::      |||   |||
Db        709 aGly--------------------------AlaLeuLeuMetHisLeuProAlaProGl 720

Qy       2842 AGCTGAGCGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTA 2901
                 |||          :::|||:::                   |||    ::::::    ::
Db        720 nAlaPheAlaLeuLeuValPhe--------------------LeuPheLysAsnTyrHi 733

Qy       2902 CAGCATGCACATTTTCCTTACTGAA-------------------------------AG 2928
              :      :::||||||    :::|||
Db        733 sLeuGlnAsnIlePheSerSerGluMetArgGlyLeuSerArgValLeuHisGlnPheTh 753

Qy       2929 GAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGCAG 2988
                 :::|||             :::      |||   :::   :::      ::::::   ::
Db        753 rArgLeuValGluAspTyrMetProSerLeuAlaIleHisPheLysArgGlnAspIleLy 773

Qy       2989 AAACTGTTCC 2998
              :   ||||||
Db        773 sThrCysSer 776
```

```
RESULT 13
NKCR_HUMAN
ID   NKCR_HUMAN     STANDARD;      PRT;   1462 AA.
AC   P30414;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE   related protein) (NK-TR protein).
GN   NKTR.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93133824; PubMed=8421688;
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS.
RA   Anderson S.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Component of a putative tumor-recognition complex.
CC       Involved in the function of NK cells.
CC   -!- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC       via its N-terminus.
CC   -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04288; AAA35734.2; -.
DR   EMBL; AF184110; AAD56402.1; -.
DR   PIR; A47328; A47328.
DR   HSSP; Q27450; 1A33.
DR   Genew; HGNC:7833; NKTR.
DR   MIM; 161565; -.
DR   GO; GO:0016018; F:cyclosporin A binding; TAS.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
FT   DOMAIN        1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN      219    240       ARG/LYS-RICH (BASIC).
FT   DOMAIN      421    457       ARG/LYS-RICH (BASIC).
FT   DOMAIN      970   1010       ARG/LYS-RICH (BASIC).
FT   DOMAIN      194    244       ARG/SER-RICH.
FT   DOMAIN      466    574       ARG/SER-RICH.
FT   DOMAIN      664    814       ARG/SER-RICH.
FT   DOMAIN     1311   1348       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE   1462 AA;  165676 MW;  D98A1147763EF527 CRC64;

Alignment Scores:
Pred. No.:                0.00689      Length:        1462
Score:                    159.00       Matches:       123
Percent Similarity:       34.34%       Conservative:  93
Best Local Similarity:    19.55%       Mismatches:    225
Query Match:              2.21%        Indels:        188
DB:                       1            Gaps:          23

US-09-302-812-3 (1-4069) x NKCR_HUMAN (1-1462)

Qy        392 CAAAAGACTATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGT 451
              |||   |||   :::  :::  |||                 :::    |||  |||
Db        675 GlnSerThrTyrSerLysTyrSerAspArgSerSerGluSerSerProArgSerArgSer 694

Qy        452 TTGGATAGTAAA------------GAAAACAACAATACAAGAATAGAATCCATGATGAGT 499
                       |||:::                 :::  ||||||     |||:::  |||
Db        695 ArgSerSerArgSerArgSerTyrSerArgSerTyrThrArgSerArgSerLeuAlaSer 714

Qy        500 TCTGTACAAAAAGAT---------AACTTTTACCAACATAATGTAGAAAAATTAGTAAAT 550
              |||        :::          :::        |||:::     |||        :::
Db        715 SerHisSerArgSerArgSerProSerSerArgSerHisSerArgAsnLysTyrSerAsp 734

Qy        551 GTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAG 610
                ||||||    |||            |||    |||     |||:::
Db        735 HisSerGlnCysSerArgSerSerSerTyrThrSerIleSerSer--------------- 749

Qy        611 CATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTG 670
                                             ::::::|||:::     ::::::   |||
Db        750 ------------------------------AspAspGlyArgArgAlaLysArgArgLeu 759

Qy        671 GAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAATGCTAAC------ 724
                 |||    :::     ::::::     ::::::    |||:::::::::
Db        760 ArgSerSerGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerGluLys 779

Qy        725 --------------------------ATTGATCGGTCACCTCAAAATGATGATCACAGT 757
                                         |||||||||                :::|||
Db        780 ThrLeuHisSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSer 799

Qy        758 GACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCA 817
              ::::::    |||      :::       :::             ::::::|||:::
Db        800 GluSerArgSerSerLeuAspTyrSerSerAspSerGluGlnSerSerValGlnAlaThr 819

Qy        818 AATGCAAAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGC 877
                 :::   |||      |||   :::               |||   ::::::|||   |||
Db        820 GlnSerAlaGlnGluLysGluLysGlnGlyGlnMetGlu---ArgThrHisAsnLysGln 838

Qy        878 AGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTG 937
                 |||      :::   ||||||:::       |||      :::   |||
Db        839 GluLys------AsnArgGlyGluGlu---LysSerLysSerGluArgGluCysProHis 855

Qy        938 GTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAAT 997
                    |||      |||   :::       |||:::   :::      |||||||||
Db        856 SerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuArgAsnGlySerLysPro 875

Qy        998 GACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAA 1057
                    |||                         |||             :::   :::
Db        876 LysArgLys--------------------------AsnTyrAlaGlySerLysTrpAsp 886

Qy       1058 AGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAA 1117
              ||||||      |||      ||||||         :::|||||||||      ||||||
Db        887 SerGluSerAsnSerGluArgAspVal------ThrLysAsnSerLysAsnAspSerHis 904

Qy       1118 GCAGATGAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACA 1177
                                   |||      |||::::::|||      ::::::|||   :::
Db        905 ------------------ProSerSerAspLysGluGluGlyGluAlaThrSerAspSer 918

Qy       1178 GCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTAC 1237
                 ::::::   |||                          :::|||    :::    :::
Db        919 GluSerGluValSer---------------------GluIleHisIleLysValLysPro 931

Qy       1238 TCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACT 1297
              :::||||||                                          :::   |||
Db        932 ThrThrLys-------------------------------------SerSerThrAsnThr 939

Qy       1298 GGAATGAATGATTTAAAT---------------GCTAAACTACCTGGAAATATTTCTAGC 1342
                 :::   |||   |||                          :::|||       :::    |||
Db        940 SerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerThrSerAsp 959

Qy       1343 CTGAATGTAGAATGCAGAAATTCTAAG---------------------------CAACAT 1375
                       |||   ||||||:::                              :::|||
Db        960 SerGluGlySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHisLysHis 979

Qy       1376 GGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGA 1435
              |||   |||::::::             :::|||          |||   :::::::::
Db        980 GlySerLysGluAsnLeuLysArgGluHis---------ThrLysLysValLysGluLys 996

Qy       1436 AGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTT 1495
                 |||   :::   :::   ||||||:::      :::||||||          :::
Db        997 LeuLysGlyLysLysAspLysLysHisLysAlaProLysArgLysGlnAlaPheHisTrp 1016

Qy       1496 CCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGA 1555
                 |||
Db       1017 GlnPro------------------------------------------------------ 1018
```

```
Qy     1556 ATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACT 1615

Db     1018 ------------------------------------------------------------ 1018

Qy     1616 ATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAA 1675
                                             |||     |||           :::
Db     1019 ---------------------------------ProLeuGluPheGlyGluGluGluGlu 1027

Qy     1676 GATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTG 1735
            :::      |||:::|||  |||             |||:::   :::
Db     1028 GluGluIleAspAspLysGlnValThrGlnGluSerLysGluLysLysVal--------- 1044

Qy     1736 GAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTT 1795
                  ||||||  |||
Db     1045 ---SerGluAsnAsnGlu------------------------------------------ 1049

Qy     1796 CTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCA 1855
                                    :::||||||  |||||||||           :::
Db     1050 -----------------------ThrIleLysAspAsnIleLeuLysThrGluLysSer 1061

Qy     1856 ------------TATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
                           |||   |||   |||           :::           |||:::
Db     1062 SerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLeuAspGln 1081

Qy     1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
                  ::::::               :::      :::   |||
Db     1082 PheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrGluGluAsn 1101

Qy     1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
            :::|||   ||||||  ||||||                          |||   :::
Db     1102 ValAla---CysLeuGlnAsnIle---------------------------GlnHisVal 1111

Qy     2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
                  |||:::               |||   :::            |||
Db     1112 GluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMetGluIle 1131

Qy     2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGT------------ 2131
            ||||||  |||   |||::::::        |||   |||   :::|||
Db     1132 CysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLeuGlyAsn 1150

Qy     2132 ------------TACCCAGACATTAAC 2146
                        |||||||||||||
Db     1151 AlaArgLeuAspThrProAspIleAsn 1159
```

RESULT 14
SPCA_HUMAN

```
ID   SPCA_HUMAN     STANDARD;      PRT;  2418 AA.
AC   P02549; Q15514;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Spectrin alpha chain, erythrocyte (Erythroid alpha-spectrin).
GN   SPTA1 OR SPTA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90170949; PubMed=1689726;
RA   Sahr K.E., Laurila P., Kotula L., Scarpa A.L., Coupal E., Leto T.L.,
RA   Linnenbach A.J., Winkelmann J.C., Speicher D.W., Marchesi V.T.,
RA   Curtis P.J., Forget B.G.;
RT   "The complete cDNA and polypeptide sequences of human erythroid
RT   alpha-spectrin.";
RL   J. Biol. Chem. 265:4434-4443(1990).
RN   [2]
RP   SEQUENCE OF 7-533 FROM N.A., AND VARIANTS EL2 PRO-260; PRO-261 AND
RP   PRO-471.
RX   MEDLINE=90009318; PubMed=2794061;
RA   Sahr K.E., Tobe T., Scarpa A., Laughinghouse K., Marchesi S.L.,
RA   Agre P., Linnenbach A.J., Marchesi V.T., Forget B.G.;
RT   "Sequence and exon-intron organization of the DNA encoding the alpha
RT   I domain of human spectrin. Application to the study of mutations
RT   causing hereditary elliptocytosis.";
RL   J. Clin. Invest. 84:1243-1252(1989).
RN   [3]
RP   SEQUENCE OF 7-601.
RX   MEDLINE=84087888; PubMed=6654896;
RA   Speicher D.W., Davis G., Marchesi V.T.;
RT   "Structure of human erythrocyte spectrin. II. The sequence of the
RT   alpha-I domain.";
RL   J. Biol. Chem. 258:14938-14947(1983).
RN   [4]
RP   SEQUENCE OF 7-125.
RX   MEDLINE=84087887; PubMed=6654895;
RA   Speicher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;
RT   "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I
RT   domain and its cyanogen bromide peptides.";
RL   J. Biol. Chem. 258:14931-14937(1983).
RN   [5]
RP   SEQUENCE OF 320-450 FROM N.A.
RX   MEDLINE=86205962; PubMed=3458204;
RA   Linnenbach A.J., Speicher D.W., Marchesi V.T., Forget B.G.;
RT   "Cloning of a portion of the chromosomal gene for human erythrocyte
RT   alpha-spectrin by using a synthetic gene fragment.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).
RN   [6]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=84295638; PubMed=6472478;
RA   Speicher D.W., Marchesi V.T.;
RT   "Erythrocyte spectrin is comprised of many homologous triple helical
RT   segments.";
RL   Nature 311:177-180(1984).
RN   [7]
RP   IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.
RA   Gibson T.J.;
RL   Unpublished observations (MAR-1995).
RN   [8]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97001215; PubMed=8844207;
RA   Maillet P., Alloisio N., Morle L., Delaunay J.;
RT   "Spectrin mutations in hereditary elliptocytosis and hereditary
RT   spherocytosis.";
RL   Hum. Mutat. 8:97-107(1996).
RN   [9]
RP   VARIANT EL2 SER-24.
RX   MEDLINE=94289716; PubMed=8018926;
RA   Parquet N., Devaux I., Boulanger L., Galand C., Boivin P.,
RA   Lecomte M.-C., Dhermy D., Garbarz M.;
RT   "Identification of three novel spectrin alpha I/74 mutations in
RT   hereditary elliptocytosis: further support for a triple-stranded
RT   folding unit model of the spectrin heterodimer contact site.";
RL   Blood 84:303-308(1994).
RN   [10]
RP   VARIANTS EL2 CYS-28; HIS-28; LEU-28 AND SER-28.
RX   MEDLINE=91358728; PubMed=1679439;
RA   Coetzer T.L., Sahr K., Prchal J., Blacklock H., Peterson L., Koler R.,
RA   Doyle J., Manaster J., Palek J.;
RT   "Four different mutations in codon 28 of alpha spectrin are
RT   associated with structurally and functionally abnormal spectrin alpha
RT   I/74 in hereditary elliptocytosis.";
RL   J. Clin. Invest. 88:743-749(1991).
RN   [11]
RP   VARIANT EL2 SER-28, AND VARIANT HPP ARG-48.
RX   MEDLINE=91346849; PubMed=1878597;
RA   Floyd P.B., Gallagher P.G., Valentino L.A., Davis M., Marchesi S.L.,
RA   Forget B.G.;
RT   "Heterogeneity of the molecular basis of hereditary
RT   pyropoikilocytosis and hereditary elliptocytosis associated with
RT   increased levels of the spectrin alpha I/74-kilodalton tryptic
RT   peptide.";
RL   Blood 78:1364-1372(1991).
RN   [12]
RP   VARIANT EL2 SER-45.
```

```
RX   MEDLINE=89323468; PubMed=2568862;
RA   Lecomte M.-C., Garbarz M., Grandchamp B., Feo C., Gautero H.,
RA   Devaux I., Bournier O., Galand C., D'Auriol L., Galibert F.,
RA   Sahr K.E., Forget B.G., Boivin P., Dhermy D.;
RT   "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white
RT   kindred with HE and HPP phenotypes.";
RL   Blood 74:1126-1133(1989).
RN   [13]
RP   VARIANT EL2/HPP PRO-207.
RX   MEDLINE=92176375; PubMed=1541680;
RA   Gallagher P.G., Tse W.T., Coetzer T., Lecomte M.-C., Garbarz M.,
RA   Zarkowsky H.S., Baruchel A., Ballas S.K., Dhermy D., Palek J.,
RA   Forget B.G.;
RT   "A common type of the spectrin alpha I 46-50a-kD peptide abnormality
RT   in hereditary elliptocytosis and pyropoikilocytosis is associated
RT   with a mutation distant from the proteolytic cleavage site. Evidence
RT   for the functional importance of the triple helical model of
RT   spectrin.";
RL   J. Clin. Invest. 89:892-898(1992).
RN   [14]
RP   VARIANT VAL-1857.
RX   MEDLINE=93253053; PubMed=8486776;
RA   Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,
RA   Kastally R., Kotula L., Delaunay J., Alloisio N.;
RT   "Low expression allele alpha LELY of red cell spectrin is associated
RT   with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and
RT   with partial skipping of exon 46.";
RL   J. Clin. Invest. 91:2091-2096(1993).
RN   [15]
RP   VARIANT EL2 BARCELONA PRO-469.
RX   MEDLINE=93372367; PubMed=8364215;
RA   dalla Venezia N., Alloisio N., Forissier A., Denoroy L., Aymerich M.,
RA   Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;
RT   "Elliptopoikilocytosis associated with the alpha 469 His-->Pro
RT   mutation in spectrin Barcelona (alpha I/50-46b).";
RL   Blood 82:1661-1665(1993).
RN   [16]
RP   VARIANT CAGLIARI GLY-2024.
RX   MEDLINE=94043025; PubMed=8226774;
RA   Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,
RA   Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,
RA   Gallanello R.;
RT   "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta
RT   spectrin repeat 17 that severely disrupts the structure and self-
RT   association of the erythrocyte spectrin heterodimer.";
RL   J. Biol. Chem. 268:22656-22662(1993).
RN   [17]
RP   VARIANT EL2 CULOZ VAL-46, AND VARIANT EL2 LYON PHE-49.
RX   MEDLINE=90347052; PubMed=2384601;
RA   Morle L., Roux A.-F., Alloisio N., Pothier B., Starck J., Denoroy J.,
RA   Morle F., Rudigoz R.-C., Forget B.G., Delaunay J., Godet J.;
RT   "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I
RT   domain. Spectrin Culoz (GGT-->GTT; alpha I 40 Gly-->Val) and
RT   spectrin Lyon (CTT-->TTT; alpha I 43 Leu-->Phe).";
RL   J. Clin. Invest. 86:548-554(1990).
RN   [18]
RP   VARIANT EL2 JENDOUBA GLU-791.
RX   MEDLINE=92345619; PubMed=1638030;
RA   Alloisio N., Wilmotte R., Morle L., Baklouti F., Marechal J.,
RA   Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,
RA   Delaunay J.;
RT   "Spectrin Jendouba: an alpha II/31 spectrin variant that is
RT   associated with elliptocytosis and carries a mutation distant from
RT   the dimer self-association site.";
RL   Blood 80:809-815(1992).
RN   [19]
RP   VARIANT EL2 TUNIS TRP-41.
RX   MEDLINE=89323436; PubMed=2568861;
RA   Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,
RA   Garbarz M., Dhermy D., Kastally R., Delaunay J.;
RT   "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due
RT   to the CGG-->TGG codon change (Arg-->Trp) at position 35 of the
RT   alpha I domain.";
RL   Blood 74:828-832(1989).
RN   [20]
RP   VARIANT EL2 GENOVA TRP-34.
RX   MEDLINE=94250920; PubMed=8193371;
RA   Perrotta S., del Giudice E.M., Alloisio N., Sciarratta G., Pinto L.,
RA   Delaunay J., Cutillo S., Lolascon A.;
RT   "Mild elliptocytosis associated with the alpha 34 Arg-->Trp mutation
RT   in spectrin Genova (alpha I/74).";
RL   Blood 83:3346-3349(1994).
RN   [21]
RP   VARIANT EL2 ANASTASIA THR-45.
RX   MEDLINE=95290423; PubMed=7772539;
RA   Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,
RA   Cutillo S., del Giudice E.M.;
RT   "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
RT   Arg-->Thr) with moderate elliptocytogenic potential.";
RL   Br. J. Haematol. 89:933-936(1995).
CC   -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal
CC       network underlying the erythrocyte plasma membrane. It associates
CC       with band 4.1 and actin to form the cytoskeletal superstructure of
CC       the erythrocyte plasma membrane.
CC   -!- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which
CC       aggregate side-to-side in an antiparallel fashion to form dimers,
CC       tetramers, and higher polymers.
CC   -!- DISEASE: Defects in SPTA1 are a cause of rhesus-unlinked
CC       elliptocytosis 2 (EL2) [MIM:130600, 182860, 166900].
CC       Elliptocytosis (also known as ovalocytosis) is a genetically
CC       heterogeneous, autosomal dominant hematologic disorder. It is

Alignment Scores:
Pred. No.:             0.00787       Length:         2418
Score:                 159.00        Matches:        131
Percent Similarity:    36.86%        Conservative:   99
Best Local Similarity: 20.99%        Mismatches:     216
Query Match:           2.21%         Indels:         178
DB:                    1             Gaps:           33

US-09-302-812-3 (1-4069) x SPCA_HUMAN (1-2418)

Qy      401 ATTACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGT 460
            :::   :::|||::::::   |||      |||   |||:::   |||       |||
Db      594 LeuLysAsnTrpIleAsnLysLysLys---LysLeuAlaAspAspGlu------AspTyr 610

Qy      461 AAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTT 520
            |||:::               |||:::::::::   |||   ||||||||
Db      611 LysAsp---------------IleGlnAsnLeuLysSerArgValGlnLysGlnGlnVal 625

Qy      521 TACCAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTA---AGTCTTGATAAGTCA 577
            ::::::      |||      ||||||   :::||||||   ::::::   |||:::
Db      626 PheGluLys------GluLeuAlaValAsnLysThrGlnLeuGluAsnIleGlnLysThr 643

Qy      578 CTCACTGAA------AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGT 631
                  |||      :::         |||   :::   :::   |||      :::
Db      644 GlyGlnGluMetIleGluGlyGlyHisTyrAlaSerAspAsnValThrThrArgLeuSer 663

Qy      632 AAG------------TGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAA 679
            :::            |||                     |||:::||||||||||:::
Db      664 GluValAlaSerLeuTrp---------------------GluGluLeuLeuGluAlaThr 676

Qy      680 CCTCAAACAGTAACCCTGGTACCAGAG---------------CAGTTTAGTAATGCTAAC 724
               |||   |||   :::   |||                  ||||||   |||
Db      677 LysGlnLysGlyThrGlnLeuHisGluAlaAsnGlnGlnLeuGlnPheGluAsn------ 694

Qy      725 ATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAATAGAGAC 784
                                                                     |||
Db      695 ---------------------------------------------------AsnAlaGluAsp 698

Qy      785 AATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAGCAGACTACGGAAGATGAA 844
               |||::::::|||      |||:::                  |||   |||   ::::::
Db      699 LeuGlnArgTrpLeuGluAspValGluTrp------------GlnValThrSerGluAsp 714

Qy      845 CACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGAC 904
            :::   :::                     |||            :::               |||
```

```
Db        715 TyrGlyLysGlyLeuAlaGluValGlnAsnArgLeuArgLysHisGlyLeuLeuGluSer 734

Qy        905 TGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTT 964
                 :::   :::||||||::::::|||:::                  |||:::|||:::
Db        735 AlaValAlaAlaArgGlnAspGlnValAspIle---------------LeuThrAspLeu 749

Qy        965 GGCTCT------GAGGATGTTGGTACTGGGTCAAAAAATGACAACAAATTGATTAGACAA 1018
                 :::      |||::::::|||                 |||     :::   ||||||
Db        750 AlaAlaTyrPheGluGluIleGlyHisProAspSerLysAspIleArg---AlaArgGln 768

Qy       1019 GAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC------------ 1066
              ||||||     :::                ||||||      ||||||
Db        769 GluSerLeuValCysArg---------PheGluAlaLeuLysGluProLeuAlaThrArg 785

Qy       1067 ---GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
                 :::      :::|||:::   :::   :::      |||:::|||:::|||    :::
Db        786 LysLysLysLeuLeuAspLeuLeuHisLeuGlnLeuIleCysArgAspThrGluAspGlu 805

Qy       1124 GAG------------GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGT--------- 1162
              |||            ||||||   |||                       |||
Db        806 GluAlaTrpIleGlnGluThrGluProSerAlaThrSerThrTyrLeuGlyLysAspLeu 825

Qy       1163 ---AGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATT 1219
                 :::|||   :::      |||:::                           |||
Db        826 IleAlaSerLysLysLeuLeuAsnArgHisArgValIleLeuGluAsnIleAlaSerHis 845

Qy       1220 GAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTA---------CATTTC 1270
              |||   |||   :::      |||:::   |||         :::         ||||||
Db        846 GluProArgIleGlnGluIleThrGluArgGlyAsnLysMetValGluGluGlyHisPhe 865

Qy       1271 CAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA 1330
                 |||         ||||||      :::      ||||||
Db        866 AlaAlaGluAspValAlaSerArg------ValLysSerLeuAsnGln------------ 879

Qy       1331 AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGAAAAAAGGATTCT 1390
              |||:::   ||||||                     :::         :::   :::
Db        880 AsnMetGluSerLeuArgAlaArgAlaAlaArgArgGlnAsnAspLeuGluAlaAsnVal 899

Qy       1391 AAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGG 1450
              :::         :::|||      |||   :::||||||         :::|||
Db        900 GlnPheGlnGlnTyrLeuAlaAspLeuHisGluAlaGluThrTrpIleArgGlu------ 917

Qy       1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCT 1510
                                         |||   |||               :::
Db        918 --------------------------LysGluProIleValAspAsnThrAsnTyrGly 928

Qy       1511 CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG-----------ATGAGAAGAATG 1558
                 |||::::::      |||      ::::::::         :::      :::
Db        929 AlaAspGluGluAlaAlaGlyAlaLeuLeuLysLysHisGluAlaPheLeuLeuAspLeu 948

Qy       1559 CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATT 1618
                       |||      :::      ||||||      ||||||
Db        949 AsnSerPheGlyAspSerMetLysAlaLeuArgAsnGlnAlaAsn--------------- 963

Qy       1619 CGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGAT 1678

Db        963 ------------------------------------------------------------ 963

Qy       1679 TTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAA 1738
                                        |||   :::|||         |||||||||
Db        964 --------------------------AlaCysGlnGlnGlnGlnAlaAlaProValGlu 974

Qy       1739 GATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTC 1798
                       ||||||:::         :::   ::::::
Db        975 GlyValAlaGlyGluGlnArgValMetAlaLeuTyrAspPhe------------------ 988

Qy       1799 AACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATAT 1858
                 :::         |||:::   :::
Db        989 ---GlnAlaArgSerProArgGluValThr------------------------------ 997

Qy       1859 TCTAAGAAATGGGACTTTACAGCTTTGATC---------------GATTTCTGGGATAAG 1903
                 ||||||   |||         |||:::               |||:::|||
Db        998 MetLysLysGlyAspValLeuThrLeuLeuSerSerIleAsnLysAspTrpTrpLys--- 1016

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
                 :::|||   |||:::                  :::|||   |||:::|||
Db       1017 ---ValGluAlaAlaAsp------------HisGlnGlyIleValPro------------ 1027

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
                 |||:::   :::      :::                  |||:::|||   |||:::
Db       1028 ---AlaValTyrValArgArgLeuAlaHisAspGluPheProMetLeuProGlnArgArg 1046

Qy       2024 AATCAT---------TCCATCACAATGTCGCAGGAACAGATT------------GCCAGT 2062
                             :::||||||      ||||||||||||                  |||
Db       1047 ArgGluGluProGlyAsnIleThrGlnArgGlnGluGlnIleGluAsnGlnTyrArgSer 1066

Qy       2063 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAG 2122
                 ||||||   |||                     |||      ::::::
Db       1067 LeuLeuAspArgAla---------------GluGluArgArgArgArgLeuLeuGlnArg 1081

Qy       2123 TATTCTAGTTAC 2134
              |||:::   :::
Db       1082 TyrAsnGluPhe 1085
```

```
RESULT 15
UN89_CAEEL
ID   UN89_CAEEL     STANDARD;      PRT;  6632 AA.
AC   O01761; Q17362;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN   UNC-89 OR C09D1.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=Bristol N2;
RX   MEDLINE=96180278; PubMed=8603916;
RA   Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT   "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT   assembly, encodes a giant modular protein composed of Ig and signal
RT   transduction domains.";
RL   J. Cell Biol. 132:835-848(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z., Le T.T., Wilson R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC       lattice assembly begins with positional cues laid down in the
CC       basement membrane and muscle cell membrane. UNC-89 responds to
CC       these signals, localizes, and then participates in assembling an
CC       M-line.
CC   -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC   -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -!- SIMILARITY: Contains 5 RCSD domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U33058; AAB00542.1; -.
DR   EMBL; AF003131; AAB54132.2; -.
DR   PDB; 1FHO; 20-DEC-00.
DR   WormPep; C09D1.1; CE30426.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR007850; RCSD.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 47.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF05177; RCSD; 5.
DR   Pfam; PF00621; RhoGEF; 1.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 23.
DR   SMART; SM00325; RhoGEF; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW   3D-structure.
FT   DOMAIN       63    127       SH3.
FT   DOMAIN      152    330       DH.
FT   DOMAIN      342    498       PH.
FT   DOMAIN      547    633       IG-LIKE C2-TYPE 1.
FT   DOMAIN      648    736       IG-LIKE C2-TYPE 2.
FT   DOMAIN      748    838       IG-LIKE C2-TYPE 3.
FT   DOMAIN      946   1033       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1044   1132       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1140   1227       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1272   1315       THR-RICH.
FT   DOMAIN     1375   1475       RCSD 1.
FT   DOMAIN     1479   1585       RCSD 2.
FT   DOMAIN     1597   1695       RCSD 3.
FT   DOMAIN     1700   1799       RCSD 4.
FT   DOMAIN     1800   1860       RCSD 5.
FT   DOMAIN     1982   2067       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2071   2163       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2171   2261       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2269   2359       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2367   2455       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2463   2564       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2563   2651       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2657   2746       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2754   2858       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2887   2980       IG-LIKE C2-TYPE 16.
FT   DOMAIN     2994   3081       IG-LIKE C2-TYPE 17.
FT   DOMAIN     3087   3183       IG-LIKE C2-TYPE 18.
FT   DOMAIN     3189   3280       IG-LIKE C2-TYPE 19.
FT   DOMAIN     3286   3376       IG-LIKE C2-TYPE 20.
FT   DOMAIN     3384   3469       IG-LIKE C2-TYPE 21.
FT   DOMAIN     3482   3572       IG-LIKE C2-TYPE 22.
FT   DOMAIN     3580   3667       IG-LIKE C2-TYPE 23.
FT   DOMAIN     3686   3777       IG-LIKE C2-TYPE 24.
FT   DOMAIN     3817   3908       IG-LIKE C2-TYPE 25.
FT   DOMAIN     3920   4009       IG-LIKE C2-TYPE 26.
FT   DOMAIN     4018   4106       IG-LIKE C2-TYPE 27.
FT   DOMAIN     4109   4201       IG-LIKE C2-TYPE 28.
FT   DOMAIN     4212   4297       IG-LIKE C2-TYPE 29.
FT   DOMAIN     4302   4387       IG-LIKE C2-TYPE 30.
FT   DOMAIN     4400   4485       IG-LIKE C2-TYPE 31.
FT   DOMAIN     4489   4580       IG-LIKE C2-TYPE 32.
FT   DOMAIN     4588   4678       IG-LIKE C2-TYPE 33.
FT   DOMAIN     4681   4771       IG-LIKE C2-TYPE 34.
FT   DOMAIN     4873   4961       IG-LIKE C2-TYPE 35.
FT   DOMAIN     4965   5057       IG-LIKE C2-TYPE 36.
FT   DOMAIN     5067   5160       IG-LIKE C2-TYPE 37.
FT   DOMAIN     5171   5260       IG-LIKE C2-TYPE 38.
FT   DOMAIN     5277   5366       IG-LIKE C2-TYPE 39.
FT   DOMAIN     5383   5472       IG-LIKE C2-TYPE 40.
FT   DOMAIN     5487   5578       IG-LIKE C2-TYPE 41.
FT   DOMAIN     5595   5685       IG-LIKE C2-TYPE 42.
FT   DOMAIN     5701   5790       IG-LIKE C2-TYPE 43.
FT   DOMAIN     5815   5904       IG-LIKE C2-TYPE 44.
FT   DOMAIN     5925   6014       IG-LIKE C2-TYPE 45.
FT   DOMAIN     6038   6130       IG-LIKE C2-TYPE 46.
FT   DOMAIN     6150   6239       IG-LIKE C2-TYPE 47.
FT   DOMAIN     6275   6368       FIBRONECTIN TYPE-III.
FT   DOMAIN     6413   6502       IG-LIKE C2-TYPE 48.
FT   DOMAIN     6507   6596       IG-LIKE C2-TYPE 49.
FT   DISULFID    568    621       POTENTIAL.
FT   DISULFID   2908   2975       POTENTIAL.
FT   DISULFID   3015   3065       POTENTIAL.
FT   DISULFID   3707   3759       POTENTIAL.
FT   DISULFID   3826   3890       POTENTIAL.
FT   DISULFID   5092   5157       POTENTIAL.
FT   DISULFID   5298   5350       POTENTIAL.
FT   DISULFID   5508   5560       POTENTIAL.
FT   DISULFID   5616   5669       POTENTIAL.
FT   DISULFID   5722   5764       POTENTIAL.
FT   DISULFID   5836   5901       POTENTIAL.
FT   DISULFID   5946   5998       POTENTIAL.
FT   DISULFID   6036   6171       POTENTIAL.
FT   DISULFID   6421   6486       POTENTIAL.
FT   CONFLICT   2137   2137       A -> P (IN REF. 1).
FT   CONFLICT   2245   2247       AKA -> PKP (IN REF. 1).
FT   CONFLICT   2258   2258       A -> P (IN REF. 1).
FT   CONFLICT   2284   2284       E -> G (IN REF. 1).
FT   CONFLICT   2297   2297       M -> I (IN REF. 1).
FT   CONFLICT   3531   3531       A -> G (IN REF. 1).
FT   CONFLICT   3884   3888       DAGEY -> RRRRI (IN REF. 1).
FT   CONFLICT   3929   3929       A -> V (IN REF. 1).
FT   CONFLICT   5134   5134       A -> P (IN REF. 1).
FT   CONFLICT   5145   5145       T -> S (IN REF. 1).
FT   CONFLICT   5185   5185       G -> A (IN REF. 1).
FT   CONFLICT   5199   5199       K -> N (IN REF. 1).
FT   CONFLICT   5202   5202       L -> F (IN REF. 1).
FT   CONFLICT   5213   5213       F -> L (IN REF. 1).
FT   CONFLICT   6178   6178       A -> G (IN REF. 1).
FT   CONFLICT   6268   6268       K -> E (IN REF. 1).
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;

Alignment Scores:
Pred. No.:              0.0111        Length:         6632
Score:                  158.50        Matches:        207
Percent Similarity:     36.18%        Conservative:   153
Best Local Similarity:  20.80%        Mismatches:     416
Query Match:            2.20%         Indels:         220
DB:                     1             Gaps:           49

US-09-302-812-3 (1-4069) x UN89_CAEEL (1-6632)

Qy        33 GCAGAGGCTGCAATCTCTGCACTTTGGGGGGCCAAGGCAGGCGCTGAGAAGGACGCGCAG 92
             :::|||:::|||   :::   :::   |||          |||      :::  :::
Db      1310 SerGluSerAlaThrThrValIleGlyGlyGlySerGlyGlyValThrGluGlySerIle 1329

Qy        93 TCCATC-----------------------------TCTCTCAGGTTAGTGAAATGAGGC 122
             |||:::                             |||        |||:::   |||
Db      1330 SerValSerLysIleGluValValSerLysThrAspSerGlnThrAspValArgGluGly 1349

Qy       123 TCTCCGCGGGGCCGGCCCGGGGACAGTGCGCTGCTGGTCCCAGCATGAATG--------- 173
             :::|||:::   |||         :::       :::|||       :::
Db      1350 ThrProLysArgArgValSerPheAlaGluGluGluLeuProLysGluValIleAspSer 1369

Qy       174 ---CGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACAACTT 230
                |||             :::|||:::|||   :::          :::|||   :::
Db      1370 AspArgLysLysLysLysSerProSerProAspLysLysGluLysSerProGluLysThr 1389

Qy       231 CGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGACCCCA 290
```

```
Db   1390 GluGluLysProAlaSerPro---ThrLysLysThrGlyGluGluValLysSer---Pro 1407

Qy    291 AGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGGCAGG 350

Db   1408 LysGluLysSerProAlaSerProThrLysLysGluLysSerProAlaAlaGluGluVal 1427

Qy    351 CGGGACAGCACAGAGGC-------AGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403

Db   1428 LysSerProThrLysLysGluLysSerProSerSerProThrLysLysGluLysSerPro 1447

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCA---------GAAAGTTTG 454

Db   1448 SerSerProThrLysLysThrGlyAspGluValLysGluLysSerProProLysSerPro 1467

Qy    455 GATAGTAAAGAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGAT 514

Db   1468 ThrLysLysGluLysSerProGluLysProGluAspValLysSerProValLysLysGlu 1487

Qy    515 AACTTTTACCAACATAATGTAGAAAAATTAGTAAATGTT---TCTCAGCTAAGTCTTGAT 571

Db   1488 Lys------SerProAspAlaThrAsnIleValGluValSerSerGluThrThrIleGlu 1505

Qy    572 AAGTCACTCACTGAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGT 631

Db   1506 LysThrGluThrThrMetThrThrGluMetThrHis------------------------ 1517

Qy    632 AAGTGGCAAAATGAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTA 691

Db   1518 ------GluSerGluGluSerArgThrSerValLysLysGluLysThrProGluLysVal 1535

Qy    692 ACCCTGGTACCAGAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGAT 751

Db   1536 AspGluLysProLys------SerProThrLysLysAspLysSerProGluLysSerIle 1553

Qy    752 CACAGTGACACAGATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAG 811

Db   1554 ThrGluGluIleLysSerProValLysLysGluLysSerProGluLysValGluGluLys 1573

Qy    812 CTTGCAAAT--------------------------------GCAAAGCAGACTACGGAA 838

Db   1574 ProAlaSerProThrLysLysGluLysSerProGluLysProAlaSerProThrLysLys 1593

Qy    839 GATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGG 898

Db   1594 SerGluAsn------GluValLysSerProThrLysLysGluLysSerProGluLysSer 1611

Qy    899 ------GAAGACTGTGCAAGTTGTCAGCAA---------GATGAGATAGACGTGGTGCCA 943

Db   1612 ValValGluGluLeuLysSerProLysGluLysSerProGluLysAlaAspAspLysPro 1631

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCT------------------GAGGATGTTGGTACT 985

Db   1632 LysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAspValLysSer 1651

Qy    986 GGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCA 1045

Db   1652 ProThrLysLysGluLysSerProGluLysValGluGlu------LysProThrSerPro 1669

Qy   1046 TTTGAGAAGGAAAGTGAACCC--------------------------------------- 1066

Db   1670 ThrLysLysGluSerSerProThrLysLysThrAspAspGluValLysSerProThrLys 1689

Qy   1067 ------GAATCACCGATGGATGTGGATAAT------------------------------ 1090

Db   1690 LysGluLysSerProGlnThrValGluGluLysProAlaSerProThrLysLysGluLys 1709

Qy   1091 ------TCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGTTTT 1144

Db   1710 SerProGluLysSerValValGluGluValLysSerProLysGluLysSerProGluLys 1729

Qy   1145 GATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGGTTCCAAGCA 1204

Db   1730 AlaGluGluLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaAla 1749

Qy   1205 AGAGACGCTGACATTGAATTTAGGAAACGGTACTCT---ACTAAGGGCGGTGAAGTTAGA 1261

Db   1750 GluGluValLysSerProThrLysLysGluLysSerProGluLysSerAlaGluGluLys 1769

Qy   1262 TTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAATGCTAAA 1321

Db   1770 ProLysSerProThrLysLysGluSerSerProValLysMetAlaAspAspGluValLys 1789

Qy   1322 CTACCT------GGAAATATTTCTAGCCTGAATGTAGAA---------TGCAGAAATTCT 1366

Db   1790 SerProThrLysLysGluLysSerProGluLysValGluGluLysProAlaSerProThr 1809

Qy   1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426

Db   1810 LysLysGluLysThrProGluLysSerAlaAlaGluGluLeuLysSerProThrLysLys 1829

Qy   1427 GAG------------------------------GACAGAAGAAAAGAACAGTGG--- 1450

Db   1830 GluLysSerProSerSerProThrLysLysThrGlyAspGluSerLysGluLysSerPro 1849

Qy   1451 GAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTTCCACCTCACCTTTCT 1510

Db   1850 GluLysProGluGluLysProLysSerProThrProLysLysSerProProGly---Ser 1868

Qy   1511 CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATGAGAAGAATGCCTCGGTGTGGG 1570

Db   1869 ProLysLysLysSerLysSerProGluAlaGluLysProProAlaProLys------ 1886

Qy   1571 ATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTT 1630

Db   1887 ------------LeuThrArgAspLeuLysLeuGlnThrValAsn---LysThrAspLeu 1901

Qy   1631 TTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAAC 1690

Db   1902 AlaHisPhe---GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGly 1920

Qy   1691 AAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGT 1750

Db   1921 LysGluIle---------ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPhe 1937

Qy   1751 GAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACA 1810

Db   1938 GluPheArgCysSer---------------IleAspThrThrMetPheGlySerGlyThr 1952

Qy   1811 CGACCCCAAAACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCT--------- 1861

Db   1953 ------------------ValSerValValAlaSerAsnAlaAlaGlySerValGluThr 1966

Qy   1862 --------------------------------------AAGAAATGGGACTTTACA--- 1879

Db   1967 LysThrGluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAsp 1986

Qy   1880 GCTTTGATCGATTTC------------TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAA 1927

Db   1987 LysLeuArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeu 2006

Qy   1928 CAT------TTATATCAG------------------------------------------ 1939

Db   2007 HisSerProLeuTyrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGly 2026

Qy   1940 ---TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAG 1996

Db   2027 ValThrIleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn------AlaGln 2044

Qy   1997 CCAATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATT 2056

Db   2045 AspSerGlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSer 2063

Qy   2057 GCCAGTCTTTTAGCTAATGCTTTCTTCTGCACA---------------TTTCCACGA--- 2098

Db   2064 AlaGlnLeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSer 2083
```

```
Qy        2099 ------------------CGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCA--- 2137
                                       |||:::   |||:::      |||   :::|||
Db        2084 ValThrIleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAla 2103

Qy        2138 ---------------GACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAA 2182
                                 ||||||      ::::::   |||         ||||||
Db        2104 ProThrValLysTrpThrIleAsn---GluLysIleValGlu------GluSerArgThr 2120

Qy        2183 CCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAA------ 2236
                        :::||||||         |||      :::::::::      |||
Db        2121 IleThrThrIleLysThrGluAspValTyrThrLeuLysIleSerAsnAlaLysIleGlu 2140

Qy        2237 CCTACTGGGTTGGTGACATTTACAAGACAGAGT---------------------CTTGAA 2275
                  ||||||.  |||      |||   |||:::
Db        2141 GlnThrGlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAla 2160

Qy        2276 GATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTACGAA 2335
               |||                     :::   |||      :::      :::|||
Db        2161 AspLeuLysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThr------ 2178

Qy        2336 GGTACCATAGAAGAAAATGGCCAAGGCATG---------CTACAGGTGGATTTTGCAAAT 2386
                     :::   :::   |||:::   :::         |||::::::|||
Db        2179 AspLysValAlaAspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp--------- 2195

Qy        2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
                                       |||   |||            |||:::   :::|||
Db        2196 ------------------GlyProSerProGly---------ThrGluValSerTrpLeu 2206

Qy        2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA--- 2503
               :::|||   :::   :::                  :::|||:::||||||   :::
Db        2207 LeuAsnGlyGlnProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGly 2226

Qy        2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGC 2548
                        :::|||      |||         |||   :::|||
Db        2227 ThrTyrHisValThrIleAlaGluAlaLysProGluMetSerGly 2241
```

Search completed: May 26, 2004, 16:52:39
Job time : 235.042 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 26, 2004, 16:13:47 ; Search time 189.952 Seconds
                                            (without alignments)
                                           13517.524 Million cell updates/sec

Title:           US-09-302-812-3
Perfect score:   7208
Sequence:        1 ggcgtctgggaagtgaggag..........agaaaaaaaaaaaaaaaaa 4069

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/6333148/runat_26052004_150053_5676/app_query.fasta_1.12437
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148_@CGN_1_1_640_@runat_26052004_150053_5676 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       SPTREMBL_25:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES

                    %
Result            Query
  No.    Score   Match Length DB  ID                        Description
---------------------------------------------------------------------------
    1     5190    72.0    976  4  Q9Y4W7                    Q9y4w7 homo sapien
    2     5113    70.9    976  4  Q86W56                    Q86w56 homo sapien
    3     5108    70.9    976  4  Q7Z742                    Q7z742 homo sapien
    4   4635.5    64.3    977  6  O02776                    O02776 bos taurus
    5     4395    61.0    968 11  O88622                    O88622 mus musculu
    6   4381.5    60.8    961 11  Q80YQ6                    Q80yq6 mus musculu
    7     4336    60.2    972 11  Q9QYM2                    Q9qym2 rattus norv
    8   4165.5    57.8    920 11  Q8CB72                    Q8cb72 mus musculu
    9     1081    15.0    768  5  O46043                    O46043 drosophila
   10     1076    14.9    723  5  Q960N8                    Q960n8 drosophila
   11      579     8.0    548 10  Q9SKB3                    Q9skb3 arabidopsis
   12      539     7.5    522 10  Q8VYA1                    Q8vya1 arabidopsis
   13      481     6.7    781  5  Q867X0                    Q867x0 caenorhabdi
   14      473     6.6    764  5  Q19637                    Q19637 caenorhabdi
   15      453     6.3    485  5  Q9N5L4                    Q9n5l4 caenorhabdi
   16    380.5     5.3    747  4  Q86V50                    Q86v50 homo sapien
   17      340     4.7    364 10  Q9SKB4                    Q9skb4 arabidopsis
   18    333.5     4.6    368  5  Q86GI4                    Q86gi4 toxoplasma
   19      188     2.6    508 12  Q9E234                    Q9e234 helicoverpa
   20    184.5     2.6   2081 10  Q9LH98                    Q9lh98 arabidopsis
   21      184     2.6    510 12  Q99GU9                    Q99gu9 helicoverpa
   22      178     2.5   1274 16  Q8NW31                    Q8nw31 staphylococ
   23      177     2.5   1163  4  Q9UHB7                    Q9uhb7 homo sapien
   24      177     2.5   1472  5  Q8I2A4                    Q8i2a4 plasmodium
   25    175.5     2.4   1078  5  Q963T1                    Q963t1 plasmodium
   26      174     2.4   1274 16  Q99TC3                    Q99tc3 staphylococ
   27    173.5     2.4    968  5  Q8I5W9                    Q8i5w9 plasmodium
   28    172.5     2.4   4969 11  Q8CF91                    Q8cf91 mus musculu
   29    172.5     2.4   5165 11  Q8CF92                    Q8cf92 mus musculu
   30      172     2.4   2241  5  Q8IK37                    Q8ik37 plasmodium
   31    171.5     2.4    838 10  Q9CAK9                    Q9cak9 arabidopsis
   32    171.5     2.4    884 10  Q9CAB4                    Q9cab4 arabidopsis
   33      171     2.4    744  5  Q86S96                    Q86s96 babesia rod
   34    170.5     2.4   5322  5  Q9VPL9                    Q9vpl9 drosophila
   35    169.5     2.4   1988  5  Q9VU16                    Q9vu16 drosophila
   36      169     2.3    498  5  Q8MTN8                    Q8mtn8 trichinella
   37      169     2.3   1745 10  Q9MAK1                    Q9mak1 arabidopsis
   38      168     2.3   1270  4  Q86XW5                    Q86xw5 homo sapien
   39      168     2.3   3167  5  Q17464                    Q17464 caenorhabdi
   40      167     2.3   4717  3  O94248                    O94248 schizosacch
   41    165.5     2.3    571  5  Q8MTN7                    Q8mtn7 trichinella
   42    165.5     2.3    596  5  Q8IE18                    Q8ie18 plasmodium
   43    165.5     2.3   1444  5  Q9VTN2                    Q9vtn2 drosophila
   44    165.5     2.3   1514  5  Q8SY55                    Q8sy55 drosophila
   45      165     2.3   5303  5  Q9V628                    Q9v628 drosophila

                                 ALIGNMENTS

RESULT 1
Q9Y4W7
ID   Q9Y4W7      PRELIMINARY;      PRT;   976 AA.
AC   Q9Y4W7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   HPARG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380098; PubMed=10449915;
RA   Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;
RT   "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT   human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT   hybridization.";
RL   Cytogenet. Cell Genet. 85:269-270(1999).
DR   EMBL; AF005043; AAB61614.1; -.
DR   Genew; HGNC:8605; PARG.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.
```

```
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   976 AA;  110903 MW;  42A0F51466FCE7DC CRC64;

Alignment Scores:
Pred. No.:              0            Length:         976
Score:                  5190.00      Matches:        976
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            72.00%       Indels:         0
DB:                     4            Gaps:           0

US-09-302-812-3 (1-4069) x Q9Y4W7 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660
```

```
Qy      2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy      2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy      2267 AGTCTTGAAGATTTTCCAGAATGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy      2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy      2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy      2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy      2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy      2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820

Qy      2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy      2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy      2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy      2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy      2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy      2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy      2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy      3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

RESULT 2
Q86W56

```
ID   Q86W56      PRELIMINARY;      PRT;   976 AA.
AC   Q86W56;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to poly (ADP-ribose) glycohydrolase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050560; AAH50560.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   976 AA;  111109 MW;  D6646353C6D0180E CRC64;

Alignment Scores:
Pred. No.:                0              Length:         976
Score:                    5113.00        Matches:        963
Percent Similarity:       98.98%         Conservative:   3
Best Local Similarity:    98.67%         Mismatches:     10
Query Match:              70.94%         Indels:         0
DB:                       4              Gaps:           0

US-09-302-812-3 (1-4069) x Q86W56 (1-976)

Qy       167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
             |||||||||||||||||||||||||||||||||      |||||||||||||||||||
Db         1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaThr 20

Qy       227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy       287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy       347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
             :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy       407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy       467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy       527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
             |||||||||||||||||   ||||||||||||||||||||||||||||||    ||||||
Db       121 HisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThrGlu 140

Qy       587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy       647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy       707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy       767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy       827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
             |||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db       221 GlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240
```

```
Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              |||   |||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        241 CysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProGlu 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              |||||||||||||||||||||||||||||||||||||||    |||||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              |||||||||||||||||||||||||||||||||||||     ||||||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440

Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||     :::||||||||||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgArgCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 3
Q7Z742
ID   Q7Z742      PRELIMINARY;      PRT;   976 AA.
AC   Q7Z742;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC052966; AAH52966.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   976 AA;  111090 MW;  47C8EE1826C4A74A CRC64;

Alignment Scores:
Pred. No.:                0              Length:         976
Score:                    5108.00        Matches:        962
Percent Similarity:       98.87%         Conservative:   3
Best Local Similarity:    98.57%         Mismatches:     11
Query Match:              70.87%         Indels:         0
DB:                       4              Gaps:           0

US-09-302-812-3 (1-4069) x Q7Z742 (1-976)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
              |||||||||||||||||||||||||||||||||||||         ||||||||||||||
Db          1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaThr 20

Qy        227 ACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTCGAC 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgArgValLeuAsp 40

Qy        287 CCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCAGGG 346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60

Qy        347 CAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATTACC 406
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80

Qy        407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100

Qy        467 AACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTACCAA 526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120

Qy        527 CATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
              ||||||||||||||||||      |||||||||||||||||||||||||||      ||||||
Db        121 HisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThrGlu 140

Qy        587 AAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAA 646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160

Qy        647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAG 706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180

Qy        707 CAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGAT 766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200

Qy        767 AGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCAAAG 826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220

Qy        827 CAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAGTCT 886
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        221 GlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

Qy        887 TGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        241 CysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProGlu 260

Qy        947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAACAAA 1006
              ||||||||||||||||||||||||||||||||||||||||||      ||||||||||||||
Db        261 SerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsnLys 280

Qy       1007 TTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAACCC 1066
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

Qy       1067 GAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGATGAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320

Qy       1127 GAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAATAAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340

Qy       1187 CCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360

Qy       1247 GGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAAT 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380

Qy       1307 GATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCT 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400

Qy       1367 AAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCA 1426
              ||||||||||||||||||||||||||||||||||||||||||      ||||||||||||||
Db        401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLysAla 420

Qy       1427 GAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCT 1486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440
```

```
Qy       1487 AAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAG 1546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460

Qy       1547 ATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCAC 1606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480

Qy       1607 ACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA 1666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500

Qy       1667 CATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAATTTG 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520

Qy       1727 TACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATTCAG 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540

Qy       1787 ACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAATAC 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560

Qy       1847 AATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580

Qy       1907 CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAAATT 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600

Qy       1967 GCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAAT 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

Qy       2027 CATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640

Qy       2087 ACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAAC 2146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660

Qy       2147 TTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTC 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

Qy       2207 TGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGACAG 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700

Qy       2267 AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800

Qy       2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
              ||||||||||||||||||||||||||||||||||||||||    :::|||   ||||||||
Db        801 TrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgHisCysThrGlu 820

Qy       2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCTGTC 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960

Qy       3047 GAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

```
RESULT 4
O02776
ID   O02776      PRELIMINARY;      PRT;   977 AA.
AC   O02776;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   BPARG.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97277328; PubMed=9115250;
RA   Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT   "Isolation and characterization of the cDNA encoding bovine poly(ADP-
RT   ribose) glycohydrolase.";
RL   J. Biol. Chem. 272:11895-11901(1997).
DR   EMBL; U78975; AAB53370.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   977 AA;  110837 MW;  87D2100F979DF377 CRC64;

Alignment Scores:
Pred. No.:              0              Length:         977
Score:                  4635.50        Matches:        874
Percent Similarity:     92.83%         Conservative:   32
Best Local Similarity:  89.55%         Mismatches:     69
Query Match:            64.31%         Indels:         1
DB:                     6              Gaps:           1

US-09-302-812-3 (1-4069) x O02776 (1-977)
```

```
Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCTACA 226
Db      1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAlaAla 20

Qy    227 ACTTCG---CCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db     21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgArgValLeu 40

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db     41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db    101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

Qy    584 GAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

Qy   1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db    601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

Qy   2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db    621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

Qy   2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db    641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

Qy   2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db    661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

Qy   2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db    681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700

Qy   2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db    701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720

Qy   2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
```

```
Db        721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||   |||||||||||||||   :::||||||   |||
Db        801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaValAlaGluArgAspValValTyr 900

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              |||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGG 3091
              ||||||||||||   :::|||   :::      |||||||||||||||
Db        961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

RESULT 5
O88622

```
ID   O88622     PRELIMINARY;      PRT;   968 AA.
AC   O88622;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   PARG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380098; PubMed=10449915;
RA   Ame J.C., Apiou F., Jacobson E.L., Jacobson M.K.;
RT   "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT   human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT   hybridization.";
RL   Cytogenet. Cell Genet. 85:269-270(1999).
DR   EMBL; AF079557; AAC28735.1; -.
DR   MGD; MGI:1347094; Parg.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   968 AA;  109169 MW;  3981DA4809E1E007 CRC64;

Alignment Scores:
Pred. No.:               0              Length:          968
Score:                   4395.00        Matches:         837
Percent Similarity:      90.99%         Conservative:    52
Best Local Similarity:   85.67%         Mismatches:      78
Query Match:             60.97%         Indels:          10
DB:                      11             Gaps:            6

US-09-302-812-3 (1-4069) x O88622 (1-968)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||   ||||||||||||||||   |||||||||||||||
Db          1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||:::::|||   |||||||:::||||||||||||   ||||||||||||||||||
Db         20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||   ||||||||:::|||||||||   ||||||||||||||||||||
Db         60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||:::|||||||||||||||||||||||||||||||||||           ||||||
Db         80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 96

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                 |||   ||||||||||||   ||||||   ||||||:::|||||||||||   |||
Db        117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              |||||||||:::||||||||||||||||   |||||||||::::::||||||||||||||
Db        137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156

Qy        644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
              ||||||||||||   ||||||||||||||||||||            |||   :::|||
Db        157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176

Qy        704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
              :::|||   |||||||||||||||   :::||||||      ||||||||||||||||||
Db        177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196

Qy        764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
              |||   ||||||:::||||||||||||||||||||||||||||   :::||||||||||||
Db        197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216

Qy        824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
              |||   |||      |||      ||||||            |||:::   |||||||||
Db        217 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 233

Qy        884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
              |||            :::||||||      ||||||||||:::|||:::||||||:::|||
Db        234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

Qy        944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
              :::|||||||||||||||||||:::||||||:::||||||||||   ||||||||||||
Db        254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273

Qy       1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
```

```
Db      274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293

Qy     1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db      294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

Qy     1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db      314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332

Qy     1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db      333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352

Qy     1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db      353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371

Qy     1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db      372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391

Qy     1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db      392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411

Qy     1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db      412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431

Qy     1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db      432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451

Qy     1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db      452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471

Qy     1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db      472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491

Qy     1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db      492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511

Qy     1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db      512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531

Qy     1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db      532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551

Qy     1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db      552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571

Qy     1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db      572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591

Qy     1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
Db      592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611

Qy     2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db      612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631

Qy     2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db      632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651

Qy     2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
Db      652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

Qy     2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
Db      672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691

Qy     2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
Db      692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

Qy     2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
Db      712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731

Qy     2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
Db      732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751

Qy     2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
Db      752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771

Qy     2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
Db      772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791

Qy     2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
Db      792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811

Qy     2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
Db      812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831

Qy     2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
Db      832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851

Qy     2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
Db      852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

Qy     2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
Db      872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891

Qy     2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
Db      892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911

Qy     2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
Db      912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931

Qy     2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
Db      932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951

Qy     3044 GTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
Db      952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 6
Q80YQ6
ID   Q80YQ6     PRELIMINARY;     PRT;   961 AA.
AC   Q80YQ6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Poly (ADP-ribose) glycohydrolase.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050892; AAH50892.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   961 AA;  108580 MW;  1558F120C367CF63 CRC64;

Alignment Scores:
Pred. No.:              0          Length:        961
Score:                  4381.50    Matches:       833
Percent Similarity:     91.12%     Conservative:  50
Best Local Similarity:  85.96%     Mismatches:    77
Query Match:            60.79%     Indels:        9
DB:                     11         Gaps:          5

US-09-302-812-3 (1-4069) x Q80YQ6 (1-961)

Qy    167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
Db      1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy    224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
Db     21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 40

Qy    284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
Db     41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60

Qy    344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
Db     61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

Qy    404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db     81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 97

Qy    464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
Db     98 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 117

Qy    524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db    118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137

Qy    584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
Db    138 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 157

Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    178 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 217

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    218 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 234

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 254

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    275 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
Db    315 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 333

Qy   1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
Db    334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353

Qy   1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
Db    354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnValGlyThr 372

Qy   1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
Db    373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392

Qy   1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
Db    393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 412

Qy   1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
Db    413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432

Qy   1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
Db    433 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 452

Qy   1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
Db    453 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 472

Qy   1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
Db    473 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 492

Qy   1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
Db    493 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 512

Qy   1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
Db    513 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 532

Qy   1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
Db    533 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 552

Qy   1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db    553 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 572

Qy   1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
Db    573 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 592
```

```
Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 612

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 632

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 652

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 672

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        673 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 692

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 712

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||||   ||||||:::||||||||||||||||||||||||
Db        713 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 732

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        733 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 752

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        753 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 772

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 792

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||||||||||:::||||||    :::|||||||||
Db        793 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 812

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 832

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        833 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 852

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 872

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 892

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db        893 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 912

Qy       2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAA 2983
              ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db        913 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 932

Qy       2984 TGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        933 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 952

Qy       3044 GTCGAGTCCTGTGCAGAGACCGCTGAC 3070
              |||||||||   |||||||||   |||
Db        953 ValGluSerSerAlaGluThrThrAsp 961
```

RESULT 7
Q9QYM2

```
ID   Q9QYM2      PRELIMINARY;      PRT;   972 AA.
AC   Q9QYM2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly(ADP-ribose) glycohydrolase.
GN   PARG.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BUF; TISSUE=Colon;
RA   Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA   Aoki Y., Nakgama H., Sugimura T.;
RT   "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RT   (Parg).";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB019366; BAA87901.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   972 AA;  109002 MW;  3DED7885684F59E9 CRC64;

Alignment Scores:
Pred. No.:                0               Length:         972
Score:                    4336.00         Matches:        822
Percent Similarity:       89.68%          Conservative:   56
Best Local Similarity:    83.96%          Mismatches:     91
Query Match:              60.16%          Indels:         10
DB:                       11              Gaps:           5

US-09-302-812-3 (1-4069) x Q9QYM2 (1-972)

Qy        167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
              |||:::||||||||||||||||||||||||||||||      ||||||||||||||   
Db          1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy        224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
              |||::::::|||   |||||||||:::|||||||||||||   ||||||:::|||||||||
Db         21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnLysArgValLeu 40

Qy        284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
              |||||||||||||||   ||||||||||||||||||||||||||||   |||||||||   
Db         41 AspProLysAspAlaProValGlnPheArgValProProSerSerSerAlaCysValSer 60

Qy        344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
              |||:::||||||   ||||||||||||   ||||||   |||||||||||||||   |||
Db         61 GlyArgAlaGlyProHisArgGlySerValThrSerPheValPheLysGlnLysProIle 80

Qy        404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
              |||:::||||||||||||||||||   ||||||||||||||||||            ||||||
Db         81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 97

Qy        464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
              ||||||||||||||||||   :::   ||||||||||||||||||||||||||||||||||
Db         98 GluAsnAsnAsnThrArgThrAspProMetMetSerSerValGlnLysAspAsnPheTyr 117

Qy        524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                 |||   ||||||||||||   ||||||   ||||||:::||||||||||||   |||
Db        118 ProHisLysValGluLysLeuGlyAsnValProGlnLeuAsnLeuAspLysSerProThr 137

Qy        584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
              ||||||||||||   ||||||||||||   ||||||||||   :::|||||||||   :::
Db        138 GluLysSerThrProTyrLeuAsnGlnGlnGlnThrAlaGlyValCysLysTrpHisSer 157
```

```
Qy    644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
Db    158 AlaGlyGluArgAlaGluGlnLeuSerAlaSerGluProSerAlaValThrGlnAlaPro 177

Qy    704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
Db    178 LysGlnLeuSerAsnAlaAsnIleAspGlnSerProProThrAspGlyHisSerAspThr 197

Qy    764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
Db    198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProValLysLeuAlaAsnAla 217

Qy    824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
Db    218 LysGlnThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerAla 234

Qy    884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
Db    235 SerCysGlnCysGlyGlnAspCysAlaGlyCysGlnArgGluGluAlaAspValIlePro 254

Qy    944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
Db    255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlySerLysAsnAspAsn 274

Qy   1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
Db    275 LysLeuThrGlyGlnGluSerGlyLeuGlyAspSerProProPheGluLysGluSerGlu 294

Qy   1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
Db    295 ProGluSerProMetAspValAspAsnSerLysThrSerCysGlnAspSerGluAlaAsp 314

Qy   1124 GAGGAGACAAGTCCAGGTTTTGATGAACAA------GAAGATGGTAGTTCCTCCCAAACA 1177
Db    315 GluGluAlaSerProValPheAspGluGlnAspAspGlnAspAspArgSerSerGlnThr 334

Qy   1178 GCAAATAAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTAC 1237
Db    335 AlaAsnLysLeuSerSerArgGlnAlaArgGluValAspGlyAspLeuArgLysArgTyr 354

Qy   1238 TCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACT 1297
Db    355 LeuThrLysGlySerGluIleArgLeuHisPheGlnPheGluGlyGly---SerAsnAla 373

Qy   1298 GGAATGAATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGC 1357
Db    374 GlyThrSerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValAspGly 393

Qy   1358 AGAAATTCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTG 1417
Db    394 ArgSerSerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheValArgIle 413

Qy   1418 CCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGG 1477
Db    414 ProLysSerGluAspLysArgLysGluGlnCysGluValArgHisGlnArgAlaGluArg 433

Qy   1478 AAGATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCC 1537
Db    434 LysIleProLysTyrValProProAsnLeuProProAspLysLysTrpLeuGlyThrPro 453

Qy   1538 ATTGAGGAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
Db    454 IleGluGluMetArgLysMetProArgCysGlyValArgLeuProLeuLeuArgProSer 473

Qy   1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657
Db    474 AlaSerHisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysPro 493

Qy   1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
Db    494 PheProThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGlu 513

Qy   1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
Db    514 GlnAsnLeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGlu 533

Qy   1778 CTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATT 1837
Db    534 LeuIleGlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIle 553

Qy   1838 CTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGG 1897
Db    554 LeuLysTyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrp 573

Qy   1898 GATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATG 1957
Db    574 AspLysValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMet 593

Qy   1958 GTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAG 2017
Db    594 ValLysIleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGln 613

Qy   2018 AAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCT 2077
Db    614 LysMetAsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAla 633

Qy   2078 TTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCA 2137
Db    634 PhePheCysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrPro 653

Qy   2138 GACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAA 2197
Db    654 AspIleAsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLys 673

Qy   2198 ACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTT 2257
Db    674 ThrLeuPheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPhe 693

Qy   2258 ACAAGACAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGA 2317
Db    694 ThrArgGlnSerLeuGluAspPheProGluTrpGluArgCysAspLysProLeuThrArg 713

Qy   2318 TTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGAT 2377
Db    714 LeuHisValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAsp 733

Qy   2378 TTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATC 2437
Db    734 PheAlaAsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIle 753

Qy   2438 CGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCAC 2497
Db    754 ArgPheLeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHis 773

Qy   2498 AATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAG 2557
Db    774 AsnGluCysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGlu 793

Qy   2558 ACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGC 2617
Db    794 ThrTyrArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCys 813

Qy   2618 TGCACTGAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTG 2677
Db    814 CysThrGluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheVal 833

Qy   2678 CCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTT 2737
Db    834 ProGluLysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyVal 853

Qy   2738 TCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGT 2797
Db    854 ProProGluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGly 873

Qy   2798 GATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTG 2857
```

```
                 |||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db          874 AspAlaArgLeuLysAlaLeuIleGlnLeuLeuAlaAlaAlaAlaAlaGluArgAspVal 893

Qy         2858 GTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTC 2917
                |||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db          894 ValTyrPheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPhe 913

Qy         2918 CTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAAT 2977
                ||||||||||||||||     ||||||   ||||||:::||||||||||||||||
Db          914 LeuThrGluArgLysLeuAsnValGlyLysValTyrArgLeuLeuLeuArgTyrTyrArg 933

Qy         2978 GAAGAATGCAGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTCATATAC 3037
                ||||||||||:::||||||:::||||||||||||   ||||||||||||||||||||
Db          934 GluGluCysArgAspCysSerSerProGlyProAspThrLysLeuTyrProPheIleTyr 953

Qy         3038 CATGCTGTCGAGTCCTGTGCAGAGACCGCTGACCATTCAGGGCAAAGGACAGGGACC 3094
                |||||   ||||||   |||||||||:::|||      ||||||||||||||||||
Db          954 HisAlaAlaGluSerSerAlaGluThrSerAspGlnProGlyGlnArgThrGlyThr 972

RESULT 8
Q8CB72
ID   Q8CB72      PRELIMINARY;      PRT;   920 AA.
AC   Q8CB72;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Poly.
GN   PARG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Bone;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK036656; BAC29519.1; -.
DR   MGD; MGI:1347094; Parg.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   920 AA;  103799 MW;  379E779E121B09A1 CRC64;

Alignment Scores:
Pred. No.:               0            Length:         920
Score:                   4165.50      Matches:        793
Percent Similarity:      90.95%       Conservative:   51
Best Local Similarity:   85.45%       Mismatches:     75
Query Match:             57.79%       Indels:         9
DB:                      11           Gaps:           5

US-09-302-812-3 (1-4069) x Q8CB72 (1-920)

Qy          167 ATGAATGCGGGCCCCGGCTGTGAACCCTGCACCAAAGCGACCCGCTGGGGCGCCGCT--- 223
                |||:::||||||||||||   ||||||||||||||         ||||||||||||||
Db            1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20

Qy          224 ACAACTTCGCCGGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCAGAGGCGCGTCCTC 283
                |||::::::|||   ||||||||:::||||||||||   ||||||||||||||||
Db           21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 40

Qy          284 GACCCCAAGGACGCTCACGTGCAGTTCAGGGTCCCACCGTCCTCGCCAGCCTGCGTCCCA 343
                ||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db           41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60

Qy          344 GGGCAGGCGGGACAGCACAGAGGCAGCGCCACCTCGCTTGTTTTCAAACAAAAGACTATT 403
                |||:::||||||   |||||||||:::||||||||   ||||||||||||||||||||
Db           61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

Qy          404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
                |||:::||||||||||||||||   ||||||||||||||||          ||||||
Db           81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu---------SerLys 97

Qy          464 GAAAACAACAATACAAGAATAGAATCCATGATGAGTTCTGTACAAAAAGATAACTTTTAC 523
                ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db           98 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 117

Qy          524 CAACATAATGTAGAAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
                   |||   |||||||||||   ||||||   ||||||:::|||||||||||   |||
Db          118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137

Qy          584 GAAAAAAGTACACAGTATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643
                |||||||||:::|||||||||||||   ||||||||::::::|||||||||||||||
Db          138 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 157

Qy          644 GAAGGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCA 703
                ||||||||||||   ||||||||||||   |||||||||         |||   :::|||
Db          158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177

Qy          704 GAGCAGTTTAGTAATGCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACA 763
                :::|||   |||||||||||||||      :::||||||      ||||||||||||||||
Db          178 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197

Qy          764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACTGTAAAGCTTGCAAATGCA 823
                |||   ||||||:::||||||||||||||||||||||||      :::||||||||||||
Db          198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 217

Qy          824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAAGCCACCAGAAGTGCAGCAAG 883
                |||   |||      |||      ||||||         |||:::   |||||||||
Db          218 LysProThrValGlyAspGlyGlnAlaArg---------SerAsnCysLysCysSerGly 234

Qy          884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAAGATGAGATAGACGTGGTGCCA 943
                |||         :::||||||      |||||||||:::|||:::||||||:::|||
Db          235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 254

Qy          944 AAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAAATGACAAC 1003
                :::||||||||||||||||||||:::||||||:::|||||||||      ||||||||||||
Db          255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274

Qy         1004 AAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCTCCATTTGAGAAGGAAAGTGAA 1063
                ||||||      |||||||||      ||||||:::||||||||||||||||||||||||||
Db          275 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294

Qy         1064 CCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314

Qy         1124 GAGGAGACAAGTCCAGGTTTTGATGAACAAGAAGATGGTAGTTCCTCCCAAACAGCAAAT 1183
                ||||||||||||||||   |||||||||||||:::|||         ||||||||||||||||
Db          315 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 333

Qy         1184 AAACCTTCAAGGTTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACT 1243
                |||   |||      |||||||||:::||||||      :::    ||||||||||||   |||
Db          334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353

Qy         1244 AAGGGCGGTGAAGTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATG 1303
                ||||||   |||||||||||||||||||||||||||   ||||||:::      |||
Db          354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 372

Qy         1304 AATGATTTAAATGCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363
                :::|||||||||||||||      ||||||   ||||||||||||||||||||||||:::
Db          373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392

Qy         1364 TCTAAGCAACATGGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAA 1423
                ||||||||||||||||||:::||||||||||||||||||||||||   ||||||:::   |||
Db          393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 412

Qy         1424 GCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATC 1483
                :::|||||||||||||||||||||      |||      :::||||||||||||||||||||||
Db          413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432

Qy         1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATTGAG 1543
```

```
                  ||||||||:::||||||:::|||    |||:::||||||||||||||||||||||||
Db        433 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 452

Qy       1544 GAGATGAGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAAT 1603
              ||||||||:::||||||||||||||||||    ||||||   |||||||||||||||:::
Db        453 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 472

Qy       1604 CACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCA 1663
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        473 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 492

Qy       1664 ACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAACAAAAT 1723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 512

Qy       1724 TTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAGCTCATT 1783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 532

Qy       1784 CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCTATTCTGAAA 1843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 552

Qy       1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        553 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 572

Qy       1904 GTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGATATGGTGAAA 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        573 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 592

Qy       1964 ATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAACAGAAGATG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 612

Qy       2024 AATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2083
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 632

Qy       2084 TGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATT 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 652

Qy       2144 AACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTC 2203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 672

Qy       2204 TTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        673 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 692

Qy       2264 CAGAGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACACGATTGCAT 2323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 712

Qy       2324 GTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCA 2383
              ||||||||||||||||||||||||    ||||||:::|||||||||||||||||||||||
Db        713 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 732

Qy       2384 AATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTT 2443
              ||||||||||||||||||||||||||||     |||||||||||||||||||||||||||
Db        733 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 752

Qy       2444 TTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2503
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        753 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 772

Qy       2504 TGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 792

Qy       2564 CGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACT 2623
              ||||||:::||||||||||||||||||||||||:::||||||     :::||||||||||
Db        793 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 812

Qy       2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAG 2683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 832

Qy       2684 AAAATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCA 2743
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        833 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 852

Qy       2744 GAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCC 2803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspVal 872

Qy       2804 AGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTAT 2863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 892

Qy       2864 TTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACT 2923
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||||||
Db        893 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 912

Qy       2924 GAAAGGAAACTCACTGTTGGAGAT 2947
              ||||||||||||    ||||||:::
Db        913 GluArgLysLeuAspValGlyGlu 920
```

RESULT 9
O46043
ID O46043 PRELIMINARY; PRT; 768 AA.
AC O46043;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ame J.-C., Jacobson M.K.;
RT   "Isolation and characterization of the cDNA encoding Drosophila
RT   poly(ADP-ribose) glycohydrolase.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 46-768 FROM N.A.
RA   Murphy L., Harris D., Barrell B.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 46-768 FROM N.A.
RA   Benos P.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003428; AAF45886.1; -.
DR   EMBL; AF079556; AAC28734.1; -.
DR   EMBL; Z98254; CAB10913.1; -.
DR   FlyBase; FBgn0023216; Parg.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   768 AA;  86288 MW;  2E620C42E643FC04 CRC64;

Alignment Scores:
Pred. No.:              2.84e-77        Length:         768
Score:                  1081.00         Matches:        246
Percent Similarity:     57.89%          Conservative:   99
Best Local Similarity:  41.28%          Mismatches:     197
Query Match:            15.00%          Indels:         54
DB:                     5               Gaps:           16

US-09-302-812-3 (1-4069) x O46043 (1-768)

Qy   1412 AGACTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACA 1471
Db     45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu------ 62

Qy   1472 GAAAGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGG 1525
Db     63 ------------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrp 74

Qy   1526 CTTGGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTG 1579
Db     75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94

Qy   1580 CCTCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCA 1639
Db     95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114

Qy   1640 GGAGAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTT 1699
Db    115 ------ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisVal 130

Qy   1700 AAAATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACT 1759
Db    131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150

Qy   1760 GCGGGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAA 1819
Db    151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170

Qy   1820 AACTTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879
Db    171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190

Qy   1880 GCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAG 1939
Db    191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210

Qy   1940 TCCATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCA 1999
Db    211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230

Qy   2000 ATACCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCC 2059
Db    231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSer 250

Qy   2060 AGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAA 2116
Db    251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270

Qy   2117 TCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCA 2176
Db    271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290

Qy   2177 AGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAG 2230
Db    291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310

Qy   2231 AAAAAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTT 2272
Db    311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330

Qy   2273 GAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTC 2326
Db    331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350

Qy   2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
Db    351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370

Qy   2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
Db    371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390

Qy   2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
Db    391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410

Qy   2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
Db    411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430

Qy   2567 TGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAG 2626
Db    431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450

Qy   2627 ATCGTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAA 2686
Db    451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470

Qy   2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------ 2728
Db    471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490

Qy   2729 CCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCC 2788
Db    491 ProGly------------------------ValAlaThrGlyAsnTrpGlyCysGlyAla 502

Qy   2789 TTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAG 2848
Db    503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522

Qy   2849 CGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATG 2908
Db    523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMet 542
```

```
Qy       2909 CACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968
Db        543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561

Qy       2969 TACTACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------ 3010
Db        562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581

Qy       3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGAC 3070
Db        582 LysLysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAsp 598

Qy       3071 CATTCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
Db        599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614

RESULT 10
Q960N8
ID   Q960N8      PRELIMINARY;      PRT;   723 AA.
AC   Q960N8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   LD42380p.
GN   PARG OR EG:114E2.1 OR CG2864.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051955; AAK93379.1; -.
DR   FlyBase; FBgn0023216; Parg.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   723 AA;  81109 MW;  90A5BA59775D611C CRC64;

 Alignment Scores:
 Pred. No.:              6.96e-77      Length:        723
 Score:                  1076.00       Matches:       245
 Percent Similarity:     57.82%        Conservative:  99
 Best Local Similarity:  41.18%        Mismatches:    197
 Query Match:            14.93%        Indels:        54
 DB:                     5             Gaps:          16

 US-09-302-812-3 (1-4069) x Q960N8 (1-723)

Qy       1415 CTGCCCAAAGCAGAGGACAGAAGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA 1474
Db          1 MetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu--------- 17

Qy       1475 AGGAAGATCCCTAAATACGTTCCACCTCACCTT------TCTCCAGATAAGAAGTGGCTT 1528
Db         18 --------------------ProGluAsnLeuAlaAsnSerLeuAspAspSerTrpArg 30

Qy       1529 GGAACTCCCATTGAGGAGATG------AGAAGAATGCCTCGGTGTGGGATCCGGCTGCCT 1582
Db         31 GlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeuPro 50

Qy       1583 CTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGA 1642
Db         51 ProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr--- 69

Qy       1643 GAAGTTCCTAAACCTTTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAA 1702
Db         70 ---ProProArgProTyrLysSerProGlyLys------TrpAspSerGluHisValArg 86

Qy       1703 ATGCCTTGTTCAGAACAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCG 1762
Db         87 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 106

Qy       1763 GGGAGCCGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGACCCCAAAAC 1822
Db        107 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 126

Qy       1823 TTGAAGGATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCT 1882
Db        127 LeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 146

Qy       1883 TTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCC 1942
Db        147 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 166

Qy       1943 ATCTTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATA 2002
Db        167 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProVal 186

Qy       2003 CCACTCCTGAAACAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGT 2062
Db        187 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 206

Qy       2063 CTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT---AAGATGAAATCG 2119
Db        207 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 226

Qy       2120 GAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGG 2179
Db        227 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 246

Qy       2180 AAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTC------ACAGAGAAA 2233
Db        247 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 266

Qy       2234 AAA------------CCTACTGGGTTGGTGACATTTACAAGACAGAGT------CTTGAA 2275
Db        267 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 286

Qy       2276 GATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTG------ACACGATTGCATGTCACT 2329
Db        287 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 306

Qy       2330 TACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGT 2389
Db        307 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 326

Qy       2390 TTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATC 2449
Db        327 TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle 346

Qy       2450 AATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTA 2509
Db        347 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 366

Qy       2510 ATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGG 2569
Db        367 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp 386

Qy       2570 TCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGATC 2629
Db        387 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 406

Qy       2630 GTTGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATG 2689
Db        407 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 426

Qy       2690 AGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------CCT 2731
```

```
Db        427 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProPro 446

Qy       2732 GGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTT 2791
              |||                       ||||||||||||||||||||||||||||||
Db        447 Gly-----------------------ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 458

Qy       2792 GGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGA 2851
              |||||||||:::   ||||||||||||:::|||::::::        |||        |||
Db        459 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 478

Qy       2852 GATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCAC 2911
               :::    |||:::|||||||||:::   |||          |||   :::    |||
Db        479 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrp 498

Qy       2912 ATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATAC 2971
              :::                        ||||||      ::::::      :::||||||
Db        499 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 517

Qy       2972 TACAATGAAGAATGCAGAAACTGTTCCACCCCTGGACCA------------------GAC 3013
              |||:::          :::      |||:::          |||
Db        518 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 537

Qy       3014 ATCAAGCTTTATCCATTCATATACCATGCTGTCGAGTCCTGTGCAGAGACCGCTGACCAT 3073
                 |||||||||  ||||||           |||         :::      |||
Db        538 LysLysLeuTyrAspPheIle---------LysGluGluLeuLysLysValArgAspVal 554

Qy       3074 TCAGGGCAA---AGGACAGGGACCTGAGGAGCCGAGCGAATAGCA 3115
                 |||:::        :::          |||:::   |||:::|||
Db        555 ProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 569
```

RESULT 11
Q9SKB3
ID Q9SKB3 PRELIMINARY; PRT; 548 AA.
AC Q9SKB3; Q94ET7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870 OR TEJ.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Panda S., Poirier G.G., Kay S.A.;
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period
RT length of the Arabidopsis circadian oscillator.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC006533; AAD32285.2; -.
DR EMBL; AF394690; AAK72256.1; -.
DR PIR; B84726; B84726.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;

Alignment Scores:

Pred. No.:	3.01e-37	Length:	548
Score:	579.00	Matches:	146
Percent Similarity:	46.56%	Conservative:	64
Best Local Similarity:	32.37%	Mismatches:	137
Query Match:	8.03%	Indels:	104
DB:	10	Gaps:	10

US-09-302-812-3 (1-4069) x Q9SKB3 (1-548)

```
Qy       1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
              :::|||:::::::::::|||   |||::::::        :::      |||:::|||
Db         79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98

Qy       1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATT-------------------------- 1987
              :::      :::   |||      |||::::::
Db         99 LeuAlaSerLeuLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118

Qy       1988 ---------------TGCACCCAGCCAATACCACTCCTGAAACAGAAGATGAATCATTCC 2032
                                  :::   ||||||      :::
Db        119 AsnIleValSerGlyIleLysThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138

Qy       2033 ATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTT 2092
              :::   :::|||||||||   |||   :::|||||||||   :::|||||||||   |||
Db        139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158

Qy       2093 CCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAAT 2152
              |||      |||         |||                  |||   |||||||||:::
Db        159 ProAspAspAsnArgGlyAlaLys------------HisLeuProValIleAsnPheAsp 174

Qy       2153 CGATTGTTTGAG------------GGACGTTCATCAAGGAAACCGGAGAAACTTAAAACG 2200
                 ||||||                        |||      :::      |||::::::
Db        175 HisLeuPheAlaSerLeuTyrIleSerTyrSerGlnSerGlnGluSerLysIleArgCys 194

Qy       2201 CTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACATTTACA 2260
              :::      ||||||   |||                  |||   |||:::|||:::|||
Db        195 IleMetHisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGlu 214

Qy       2261 AGACAG------AGTCTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAAACCCTTGACA 2314
              |||:::      :::      |||         |||   :::   :::      |||
Db        215 ArgLysIleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeuCys 234

Qy       2315 CGATTGCATGTCACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTG 2374
                       |||         |||   ||||||:::               |||:::|||
Db        235 AlaPheLysValHisSerPheGlyLeuIleGluAspGlnProAspAsnAlaLeuGluVal 254

Qy       2375 GATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAA 2434
              ||||||||||||:::::::::|||||||||         |||   |||   ||||||||||||
Db        255 AspPheAlaAsnLysTyrLeuGlyGlyGlySerLeuSerArgGlyCysValGlnGluGlu 274

Qy       2435 ATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGAT 2494
              |||||||||:::||||||||||||||||||         ||||||         :::|||
Db        275 IleArgPheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAsp 294

Qy       2495 CACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCT 2554
                 ||||||   :::   |||   |||   |||::::::|||   ||||||||||||||
Db        295 AspAsnGluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAla 314

Qy       2555 GAGACATATCGTTGGTCCCGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGG 2614
                 ::::::|||::::::      :::   |||               |||      :::|||
Db        315 SerSerPheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspProPheLysArg 334

Qy       2615 CGCTGCACTGAGATCGTTGCCATCGATGCTCTT------------------CACTTCAGA 2656
              |||   |||   ||||||||||||||||||||||||                  ||||||:::
Db        335 ArgArgThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLys 354

Qy       2657 CGCTACCTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTACTGT 2716
                                         :::   ||||||:::|||||||||   |||
Db        355 AspIleCys---------------------LeuLeuArgGluIleAsnLysAlaLeuCys 367

Qy       2717 GGATTTCTCCGT------------------------------------------------ 2728
              |||||||||
Db        368 GlyPheLeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAsp 387

Qy       2728 ------------------------------------------------------------ 2728

Db        388 AsnGluIleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThr 407

Qy       2728 ------------------------------------------------------------ 2728
```

```
Db        408 AlaSerHisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsn 427

Qy       2729 ---------------------CCTGGAGTTTCTTCAGAGAATCTT------TCTGCAGTG 2761
                                   ||||||  :::|||:::                |||
Db        428 LeuIleArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyVal 447

Qy       2762 GCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATA 2821
              ||||||||||||||||||||||||   |||||||||||||      |||||||||   |||
Db        448 AlaThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIle 467

Qy       2822 CAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTT---TATTTCACCTTTGGGGAC 2878
               |||   |||||||||:::           |||        :::    |||:::||||||||
Db        468 GlnTrpLeuAlaAlaSerGlnThrArgArgProPheIleSerTyrTyrThrPheGly--- 486

Qy       2879 TCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACT 2938
                 |||    :::|||::::::         :::       ::::::        |||   |||
Db        487 ValGluAlaLeuArgAsnLeuAspGlnValThrLysTrpIleLeuSerHisLysTrpThr 506

Qy       2939 GTTGGAGATGTGTATAAGCTGTTGCTACGATAC 2971
               ||||||||||::::::    ::::::|||    |||
Db        507 ValGlyAspLeuTrpAsnMetMetLeuGluTyr 517

RESULT 12
Q8VYA1
ID   Q8VYA1      PRELIMINARY;      PRT;    522 AA.
AC   Q8VYA1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative poly(ADP-ribose) glycohydrolase.
GN   AT2G31870.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY072330; AAL61937.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hydrolase.
SQ   SEQUENCE   522 AA;  58942 MW;  6FF98685D867CA15 CRC64;

Alignment Scores:
Pred. No.:              4.62e-34        Length:          522
Score:                  539.00          Matches:         133
Percent Similarity:     49.65%          Conservative:    82
Best Local Similarity:  30.72%          Mismatches:      138
Query Match:            7.48%           Indels:          80
DB:                     10              Gaps:            11

US-09-302-812-3 (1-4069) x Q8VYA1 (1-522)

Qy       1892 TTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCT 1951
               |||:::||||||    :::              |||:::    :::    :::      ::::::|||
Db         80 PhePheAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValValPro 99

Qy       1952 GATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTG 2011
                 :::   ::::::      |||      ||||||::::::          :::
Db        100 AlaLeuCysArgLeuLeuLeuGlnLeuProSerMetLeuGluLysHisTyr--------- 116

Qy       2012 AAACAGAAGATGAATCAT------------------------------------------ 2029
                 ||||||    :::|||
Db        117 ---GlnLysAlaAspHisValLeuAspGlyValLysSerGlyLeuArgLeuLeuGlyPro 135

Qy       2030 ------------TCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAATGCT 2077
                         :::    :::|||||||||    ||||||:::|||||||||    :::
Db        136 GlnGluAlaGlyIleValLeuLeuSerGlnGluLeuIleAlaAlaLeuLeuAlaCysSer 155

Qy       2078 TTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTACCCA 2137
              |||||||||    ||||||          :::      :::|||                :::
Db        156 PhePheCysLeuPheProGluValAspArgSerLeuLys------------AsnLeuGln 171

Qy       2138 GACATTAACTTCAATCGATTGTTTGAG------------GGACGTTCATCAAGGAAACCG 2185
                 |||||||||:::   ||||||                            :::::::::
Db        172 GlyIleAsnPheSerGlyLeuPheSerPheProTyrMetArgHisCysThrLysGlnGlu 191

Qy       2186 GAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGG 2245
                 |||:::|||    |||          ||||||    |||:::                |||||||||
Db        192 AsnLysIleLysCysLeuIleHisTyrPheGlyArgIleCysArgTrpMetProThrGly 211

Qy       2246 TTGGTGACATTTACAAGACAGAGTCTTGAA------------------------GATTTT 2281
                 |||:::|||    |||:::    |||                              :::
Db        212 PheValSerPheGluArgLysIleLeuProLeuGluTyrHisProHisPheValSerTyr 231

Qy       2282 CCAGAA---------TGGGAAAGATGTGAAAAACCCTTGACACGATTGCATGTCACTTAC 2332
              |||:::            |||                ||||||       :::    :::
Db        232 ProLysAlaAspSerTrpAlaAsnSerValThrProLeuCysSerIleGluIleHisThr 251

Qy       2333 GAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAATCGTTTT 2392
                 |||    ||||||:::               |||:::||||||||||||:::    :::
Db        252 SerGlyAlaIleGluAspGlnProCysGluAlaLeuGluValAspPheAlaAspGluTyr 271

Qy       2393 GTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTAATCAAT 2452
                 ||||||          |||          :::||||||||||||||||||:::||||||
Db        272 PheGlyGlyLeuThrLeuSerTyrAspThrLeuGlnGluGluIleArgPheValIleAsn 291

Qy       2453 CCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGTCTAATT 2512
              ||||||||||||       :::|||          :::|||    ||||||    :::
Db        292 ProGluLeuIleAlaGlyMetIlePheLeuProArgMetAspAlaAsnGluAlaIleGlu 311

Qy       2513 ATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTTGGTCC 2572
              |||   |||   |||::::::|||   ||||||||||||          ::::::::::::::
Db        312 IleValGlyValGluArgPheSerGlyTyrThrGlyTyrGlyProSerPheGlnTyrAla 331

Qy       2573 CGGAGCCACGAAGATGGGAGTGAAAGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTT 2632
                 :::   |||      :::   |||            ||||||   |||    ::::::
Db        332 GlyAspTyrThrAspAsnLysAspLeuAspIlePheArgArgArgLysThrArgValIle 351

Qy       2633 GCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATGAGA 2692
              ||||||||||||:::                :::   |||:::       :::    :::
Db        352 AlaIleAspAlaMetPro---AspProGlyMetGlyGlnTyrLysLeuAspAlaLeuIle 370

Qy       2693 CGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGT------------------------ 2728
              ||||||:::|||||||||:::     |||::::::
Db        371 ArgGluValAsnLysAlaPheSerGlyTyrMetHisGlnCysLysTyrAsnIleAspVal 390

Qy       2729 ---------CCTGGAGTTTCTTCA------------------------------------ 2743
                          |||      ||||||
Db        391 LysHisAspProGluAlaSerSerSerHisValProLeuThrSerAspSerAlaSerGln 410

Qy       2744 ------------------------------------------GAGAATCTTTCTGCAGTGGCC 2764
                                                         |||         ||||||
Db        411 ValIleGluSerSerHisArgTrpCysIleAspHisGluGluLysLysIleGlyValAla 430

Qy       2765 ACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAG 2824
              ||||||||||||||||||||||||   ||||||||||||         ||||||    ::::::|||
Db        431 ThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysIleMetLeuGln 450

Qy       2825 ATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTT---TATTTCACCTTTGGGGACTCA 2881
                 ||||||    :::    :::   |||         :::    |||:::|||||||||
Db        451 TrpLeuAlaIleSerGlnSerGlyArgProPheMetSerTyrTyrThrPheGly---Leu 469

Qy       2882 GAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACTGTT 2941
                 :::    ::::::::::::       :::          :::         :::::::::||||||
Db        470 GlnAlaLeuGlnAsnLeuAsnGlnValIleGluMetValAlaLeuGlnGluMetThrVal 489
```

```
Qy       2942 GGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAA 2980
              ||||||::::::|||   |||:::   |||   :::|||
Db        490 GlyAspLeuTrpLysLysLeuValGluTyrSerSerGlu 502

RESULT 13
Q867X0
ID   Q867X0      PRELIMINARY;      PRT;   781 AA.
AC   Q867X0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3
DE   protein) (Corresponding sequence F20C5.1a).
GN   F20C5.1 OR PME-3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RT   Caenorhabditis elegans.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY185493; AAO26316.1; -.
DR   EMBL; Z68161; CAD89735.1; -.
DR   WormPep; F20C5.1a; CE33775.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   781 AA;  89273 MW;  30447D06605CED0A CRC64;

Alignment Scores:
Pred. No.:              2.26e-29      Length:         781
Score:                  481.00        Matches:        221
Percent Similarity:     40.82%        Conservative:   137
Best Local Similarity:  25.20%        Mismatches:     308
Query Match:            6.67%         Indels:         211
DB:                     5             Gaps:           39

US-09-302-812-3 (1-4069) x Q867X0 (1-781)

Qy        599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
              :::   :::|||      |||             |||         |||     |||
Db         27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              |||         :::                         ||||||:::      :::
Db         47 GluSerLysGluAsp-------------------------ProGluGluProLysSer 57

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
               ::::::      |||   |||:::   |||   |||      ||||||   ||||||
Db         58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy        779 --------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                  :::|||   :::   ||||||   |||:::|||:::
Db         78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy        815 GCAAATGCAAAGCAG------------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
              :::|||   :::              |||         ||||||      :::
Db         98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy        863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
              :::|||             ||||||   |||   |||:::
Db        116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy        923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
                       ||||||      ||||||   :::                  ||||||
Db        126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy        983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
                 |||      |||::::::|||:::   :::|||:::
Db        137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy       1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
                    ||||||                              :::   :::   :::
Db        151 ------AsnLysGlu-----------------------MetArgAspGlyGluGlnVal 160

Qy       1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
                 ||||||      ||||||::::::               ||||||      :::
Db        161 GlnGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 180

Qy       1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                                   |||      :::||||||   |||
Db        181 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 200

Qy       1196 TTCCAAGCAAGAGACGCTGACATTGAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAA 1255
                 :::      :::||||||                     ||||||   ||||||
Db        201 IleGluGluThrGluAlaAsp---------------------SerThrPheValGlyGlu 213

Qy       1256 GTTAGATTACATTTCCAATTTGAAGGAGGAGAGAGTCGCACTGGAATGAATGATTTAAAT 1315
                 :::   :::                  :::|||   |||      ::::::::::::
Db        214 AspSerLysAsnGlnArg------------GlnSerGlyThrThrSerAspGluValAsp 229

Qy       1316 GCTAAACTACCTGGAAATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACAT 1375
              |||         |||:::::::::            |||         :::|||
Db        230 AlaAspSerGlnIleAsnLeuAlaThrLysThrValArgThrSerSerSer--------- 246

Qy       1376 GGAAAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGA 1435
                                      |||      :::               |||
Db        247 ------------------------SerPheLeuSerThrValSerThrCysGluAlaPro 258

Qy       1436 AGAAAAGAACAGTGGGAAACCAAACATCAAAGAACAGAAAGGAAGATCCCTAAATACGTT 1495
                 |||   :::                  :::      |||:::   :::   :::
Db        259 AlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThr 278

Qy       1496 CCACCTCACCTTTCT------CCAGATAAGAAGTGGCTTGGAACTCCCATTGAGGAGATG 1549
                 :::|||:::         ||||||
Db        279 GluGlyAsnLeuThrLeuGlnProAsp--------------------------------- 287

Qy       1550 AGAAGAATGCCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCTGCCAATCACACA 1609
                                          |||      :::   |||      |||:::
Db        288 --------------------------LeuAsnLysValAspProAspArgAsnTyr--- 297

Qy       1610 GTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACACAT 1669
                                       |||      :::|||      ||||||
Db        298 ------------------------ArgTyrCysThrIleProAsn---PheProAlaSer 308

Qy       1670 TATAAAGATTTG---TGGATAACAAGCAT---------GTTAAAATGCCTTGTTCAGAA 1717
                    |||      ||||||::::::          :::      :::|||
Db        309 GlnGlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeuPro--------- 325

Qy       1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
                                                            ||||||
Db        326 ------------------------------------------------GlnArgTrpArg 329

Qy       1778 CTCATTCAGACTGCACTTCTCAACAA-----------ATTTACACGACCCCAAAACTTG 1825
                 |||   |||         ::::::               |||         ||||||:
Db        330 -GluPhe-AspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLysLeuA 349

Qy       1826 AAGGA---------TGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGA---C 1873
               ::|||         ||||||                  |||   |||      :::|||   |
Db        349 spGlyTyrLeuLysCysTyrLys------ThrThrGlyTyrPheMetPheValGlyLeuL 367

Qy       1874 TTTACAGCTTTGATCGATTTCTGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTA 1933
```

```
                 ||::::: :::    :::|||          :::      :::     |||     :::
Db        367 euHisAsn-MetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGluMet 386

Qy       1934 TATCAGTCCATCTTGCCTGATATGGTG--------------------------------- 1960
              |||       :::   ::::::|||
Db        387 TyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAlaArg 406

Qy       1961 ------AAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
                 :::||||||         :::|||     :::    |||      |||
Db        407 ValAlaArgIleAlaLysThrAlaGluAspIleLeuProGluArgIleTyrArgLeuVal 426

Qy       2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
                 :::      |||   |||:::|||    :::|||   |||:::|||:::|||
Db        427 GlyAspVal---GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 445

Qy       2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
                 ||||||                       |||:::     |||    :::|||
Db        446 MetPhePhe---------------------AlaArgProAspSerProPheSer------ 456

Qy       2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                          |||  |||:::          |||        |||||||||
Db        457 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 472

Qy       2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
              |||   ||||||   ||||||   :::::::::          |||   |||   |||:::
Db        473 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 492

Qy       2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGAAAGATGTGAAAAA 2305
              |||      |||:::        :::    |||       ||||||:::      :::|||
Db        493 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 510

Qy       2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
                 |||   |||          ::::::     :::    |||:::
Db        511 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 530

Qy       2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
                    ||||||||||||||||||      :::||||||||||||    :::     |||
Db        531 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 550

Qy       2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
                 ||||||||||||||||||||||||:::   ||||||:::::::::        |||
Db        551 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 570

Qy       2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
                 |||   :::         |||   :::    |||   |||      :::|||
Db        571 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 590

Qy       2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
              ||||||||||||      |||   :::|||                :::   |||   :::
Db        591 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 610

Qy       2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
              :::   :::|||   ||||||          |||         |||   :::|||||||||
Db        611 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 630

Qy       2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
              |||:::   |||:::                          :::|||
Db        631 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 650

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              :::    ||||||:::    ||||||         ||||||:::      |||
Db        651 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 670

Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              |||:::      :::         ||||||   |||||||||||||||||   ||||||
Db        671 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 689

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||   :::||||||::::::||||||      |||:::|||    :::     :::
Db        690 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 709

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::||||||:::   ||||||                  :::           :::  :::
Db        710 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 729

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              :::   :::|||:::|||   ::::::   ::::::
Db        730 LysAspValThrLeuGlyMetLeuPheSerMetIle------------------------ 741

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTC 3031
                 |||         |||            :::   |||   |||
Db        742 AsnAsnThrGlyLeuProHisLysHisPheGluPheTyrValPhe 756
```

RESULT 14

```
Q19637
ID   Q19637     PRELIMINARY;     PRT;   764 AA.
AC   Q19637;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose
DE   metabolism enzyme-3 short form).
GN   F20C5.1 OR PME-3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RT   Caenorhabditis elegans.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z68161; CAA92299.2; -.
DR   EMBL; AY185494; AAO26317.1; -.
DR   PIR; T21138; T21138.
DR   WormPep; F20C5.1b; CE32867.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
SQ   SEQUENCE   764 AA;  87441 MW;  B391703B1D0CB79A CRC64;

Alignment Scores:
Pred. No.:               9.77e-29      Length:         764
Score:                   473.00        Matches:        211
Percent Similarity:      37.88%        Conservative:   128
Best Local Similarity:   23.58%        Mismatches:     290
Query Match:             6.56%         Indels:         266
DB:                      5             Gaps:           38

US-09-302-812-3 (1-4069) x Q19637 (1-764)

Qy        599 TATTTGAACCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAATGAAGGGAAACACACG 658
              :::   :::|||      |||               |||               |||   |||
Db         27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46

Qy        659 GAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAACCCTGGTACCAGAGCAGTTTAGTAAT 718
              |||      :::                                 ||||||:::   :::
Db         47 GluSerLysGluAsp---------------------------ProGluGluProLysSer 57

Qy        719 GCTAACATTGATCGGTCACCTCAAAATGATGATCACAGTGACACAGATAGTGAAGAGAAT 778
                 ::::::      |||   |||:::   |||   |||      ||||||   ||||||
Db         58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77

Qy        779 --------------------AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
                                  :::|||   :::   ||||||   |||:::|||:::
```

```
Db         78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97

Qy        815 GCAAATGCAAAGCAG-----------ACTACGGAAGATGAACACGCCAGAGAAGCCAAA 862
              :::|||   :::             |||         ||||||       :::
Db         98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp------ 115

Qy        863 AGCCACCAGAAGTGCAGCAAGTCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTCAGCAA 922
              :::|||            ||||||        |||   |||:::
Db        116 AsnHis------------LysSerThrGluProMetGluGlu------------------ 125

Qy        923 GATGAGATAGACGTGGTGCCAAAGAGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGT 982
                       ||||||   ||||||   :::                  ||||||
Db        126 ---------AspValAsnAsnLysSerAsnIle------------------AspValAla 136

Qy        983 ACTGGGTCAAAAAATGACAACAAATTGATTAGACAAGAAAGTTGCCTAGGAAATTCTCCT 1042
                    |||      |||::::::|||:::    :::|||:::
Db        137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn------------------ 150

Qy       1043 CCATTTGAGAAGGAAAGTGAACCCGAATCACCGATGGATGTGGATAATTCTAAAAATAGT 1102
                    ||||||                   |||   :::   :::
Db        151 ------AsnLysGluMetArg---------------AspGlyGluGlnValGlnGlnLeu 163

Qy       1103 TGTCAAGACTCAGAAGCAGATGAGGAG---------ACAAGTCCAGGTTTTGATGAACAA 1153
                ||||||     ||||||:::::          ||||||       :::
Db        164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183

Qy       1154 GAA------------------GATGGTAGTTCCTCCCAAACAGCAAATAAACCTTCAAGG 1195
                                   |||    :::||||||   |||
Db        184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203

Qy       1196 TTCCAAGCAAGAGACGCTGACATTGAATTT---------------------AGGAAACGG 1234
                :::      :::||||||      |||                      |||
Db        204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223

Qy       1235 TACTCTACTAAGGGC--------------------------------------------- 1249
                   |||
Db        224 ValArgThrSerSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243

Qy       1250 GGTGAAGTTAGATTACAT-----------------------TTCCAATTTGAAGGAGGA 1285
              |||      |||:::::                        |||       |||
Db        244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263

Qy       1286 GAGAGTCGCACTGGAATGAATGATTTAAATGCTAAACTACCTGGA--------------- 1330
                 :::   |||         |||||||||       |||
Db        264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282

Qy       1331 ------------AATATTTCTAGCCTGAATGTAGAATGCAGAAATTCTAAGCAACATGGA 1378
                          |||         :::         :::   |||          ::::::|||
Db        283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302

Qy       1379 AAAAAGGATTCTAAAATCACAGATCATTTGATGAGACTGCCCAAAGCAGAGGACAGAAGA 1438
                 |||                         :::   ||||||          :::|||
Db        303 ProLys------------------------IleValLeuPro------------GlnArg 310

Qy       1439 AAAGAACAGTGGGAAACCAAACATCAAAGAACAGAA------------------AGGAAG 1480
                 ::::::::::::::::     :::|||    :::               ||||||
Db        311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330

Qy       1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCATT 1540
              :::         |||:::         :::   :::                     :::
Db        331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350

Qy       1541 GAGGAGATGAGAAGAATG---CCTCGGTGTGGGATCCGGCTGCCTCTCTTGAGACCATCT 1597
                    |||         |||         :::||||||   |||
Db        351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu------ 368

Qy       1598 GCCAATCACACAGTAACTATTCGGGTAGATCTTTTGCGAGCAGGAGAAGTTCCTAAACCT 1657

Db        368 ------------------------------------------------------------ 368

Qy       1658 TTTCCAACACATTATAAAGATTTGTGGGATAACAAGCATGTTAAAATGCCTTGTTCAGAA 1717
                    :::||||||::::::
Db        369 ------MetTyrTyrLysGluMet------------------------------------ 374

Qy       1718 CAAAATTTGTACCCAGTGGAAGATGAGAATGGTGAGCGAACTGCGGGGAGCCGGTGGGAG 1777
                                                                      |||
Db        375 ------------------------------------------------------SerGlu 376

Qy       1778 CTCATT---CAGACTGCACTTCTCAACAAATTTACACGACCCCAAAACTTGAAGGATGCT 1834
              |||:::   :::      :::|||   ||||||   |||                :::
Db        377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla------ 394

Qy       1835 ATTCTGAAATACAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATCGATTTC 1894

Db        394 ------------------------------------------------------------ 394

Qy       1895 TGGGATAAGGTACTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTTGCCTGAT 1954
                    |||      |||:::          :::   :::|||:::   ::::::   |||
Db        395 ------LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411

Qy       1955 ATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCACTCCTGAAA 2014
              :::
Db        412 Val--------------------------------------------------------- 412

Qy       2015 CAGAAGATGAATCATTCCATCACAATGTCGCAGGAACAGATTGCCAGTCTTTTAGCTAAT 2074
                          |||   |||:::|||   :::|||   |||:::|||:::|||
Db        413 ------------GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428

Qy       2075 GCTTTCTTCTGCACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTATTCTAGTTAC 2134
                ||||||                     |||:::      |||   :::|||
Db        429 MetPhePhe--------------------AlaArgProAspSerProPheSer------ 439

Qy       2135 CCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTCATCAAGGAAACCGGAGAAACTT 2194
                          |||   |||:::                  |||         ||||||||
Db        440 ------------PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455

Qy       2195 AAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAAACCTACTGGGTTGGTGACA 2254
              |||   ||||||   ||||||   :::::::::               |||   |||   |||:::
Db        456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer 475

Qy       2255 TTT------ACAAGACAGAGTCTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAAAA 2305
              |||      |||:::         :::   |||      ||||||:::         :::|||
Db        476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTrpLys------AspLys 493

Qy       2306 CCCTTGACACGATTGCATGTCACT------TACGAAGGTACCATAGAAGAAAATGGCCAA 2359
                 |||      |||               ::::::      :::   |||:::
Db        494 LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla 513

Qy       2360 GGCATGCTACAGGTGGATTTTGCAAATCGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGA 2419
                       ||||||||||||||||         :::||||||||||||   :::   |||
Db        514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGly 533

Qy       2420 CTTGTGCAAGAAGAAATCCGCTTTTTAATCAATCCTGAGTTGATTATTTCACGGCTCTTC 2479
                 ||||||||||||||||||||||:::   ||||||:::::::::         |||
Db        534 SerValGlnGluGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu 553

Qy       2480 ACTGAGGTGCTGGATCACAATGAATGTCTAATTATCACAGGTACTGAGCAGTACAGTGAA 2539
                 |||   :::         |||   :::   |||   |||            :::|||
Db        554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573

Qy       2540 TACACAGGCTATGCTGAGACATATCGTTGG---------------TCCCGGAGCCACGAA 2584
              ||||||||||||      |||   :::|||                  :::   |||   :::
Db        574 TyrThrGlyTyrGlyHisThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGln 593

Qy       2585 GATGGGAGTGAA---AGGGACGACTGCGAGCGGCGCTGCACTGAGATCGTTGCCATCGAT 2641
              :::   :::|||   ||||||         |||         |||   :::|||||||||
Db        594 AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp 613

Qy       2642 GCTCTTCACTTCAGACGC---------------TACCTCGATCAGTTTGTGCCTGAGAAA 2686
              |||:::   |||:::                           :::|||
Db        614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633

Qy       2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAG 2746
              :::   ||||||:::   ||||||      ||||||:::         |||
Db        634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653
```

```
Qy       2747 AATCTTTCTGCAGTGGCCACAGGAAACTGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG 2806
              |||:::     :::  ||||||  ||||||||||||||||||    ||||||
Db        654 AsnIle---ProIleValThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysPro 672

Qy       2807 TTAAAAGCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCGAGATGTGGTTTATTTC 2866
              ||||||   :::||||||::::::||||||        |||:::|||   :::   :::
Db        673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692

Qy       2867 ACCTTTGGGGACTCAGAATTGATGAGAGACATTTACAGCATGCACATTTTCCTTACTGAA 2926
              :::||||||:::   ||||||                :::            :::   :::
Db        693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712

Qy       2927 AGGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGATACTACAATGAAGAATGC 2986
              :::   :::|||:::|||   ::::::   ::::::
Db        713 LysAspValThrLeuGlyMetLeuPheSerMetIle------------------------ 724

Qy       2987 AGAAACTGTTCCACCCCTGGACCAGACATCAAGCTTTATCCATTC 3031
                 |||      |||            :::   |||   |||
Db        725 AsnAsnThrGlyLeuProHisLysHisPheGluPheTyrValPhe 739
```

```
RESULT 15
Q9N5L4
ID   Q9N5L4      PRELIMINARY;      PRT;   485 AA.
AC   Q9N5L4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme
DE   4).
GN   H23L24.5 OR PME-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Abu-Threideh J., Lehnert L.;
RT   "The sequence of C. elegans cosmid H23L24.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Gagnon S.N., Hardy I., Desnoyers S.;
RT   "Molecular characterization of poly ADP-ribose glycohydrolases in
RT   Caenorhabditis elegans.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006662; AAF39896.2; -.
DR   EMBL; AF548468; AAN40699.1; -.
DR   WormPep; H23L24.5; CE32685.
DR   InterPro; IPR007724; PARG.
DR   Pfam; PF05028; PARG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   485 AA;  55574 MW;  86367DDBAB15081F CRC64;

Alignment Scores:
Pred. No.:              3.3e-27        Length:         485
Score:                  453.00         Matches:        136
Percent Similarity:     47.40%         Conservative:   74
Best Local Similarity:  30.70%         Mismatches:     157
Query Match:            6.28%          Indels:         76
DB:                     5              Gaps:           15

US-09-302-812-3 (1-4069) x Q9N5L4 (1-485)
```

```
Qy       1772 TGGGAGCTCATTCAGACTGCACTTCTCAACAAATTTACACGA---CCCCAAAACTTGAAG 1828
              :::||||||::::::|||   :::      ::::::                 ||||||||
Db         46 PheGluLeuLeuGluThrGlyValSerGlnGlnTrpArgAsnCysAspGlnAsnLeuPhe 65

Qy       1829 GATGCTATTCTGAAA------TACAATGTGGCATATTCT-------------------- 1861
              :::      ||||||         |||      ||||||
Db         66 AsnGluTyrLeuLysThrTyrLysAsnGlyGlyTyrSerGlnPheGluAspLeuLeuPhe 85

Qy       1862 --------------------------AAGAAATGGGACTTTACAGCTTTGATCGATTTC 1894
                                        :::::::::|||      ||||||       |||
Db         86 LysIleTrpGlyTyrSerGluGluLysGluArgPheAspLeuProAlaLeuLysSerPhe 105

Qy       1895 TGGGATAAGGTACTTGAA---------GAAGCAGAAGCTCAACATTTATATCAGTCCATC 1945
              :::   |||:::   |||         |||   |||                     :::
Db        106 TyrArgLysMetSerGluIleValGlyGluAspGlu---------------------Val 118

Qy       1946 TTGCCTGATATGGTGAAAATTGCACTCTGTCTGCCAAATATTTGCACCCAGCCAATACCA 2005
              |||      :::   ::::::                  :::   |||
Db        119 LeuGluLysLeuAlaArgLeuValArgIleThrLysSerAlaCys------------Glu 134

Qy       2006 CTCCTGAAACAGAAGATGAAT------------------CATTCCATCACAATGTCGCAG 2047
              :::|||   :::|||:::                     |||   |||   |||
Db        135 ValLeuProGluLysIleTyrArgLeuValGlyAspIleGluSerAlaThrPheSerHis 154

Qy       2048 GAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCACGACGAAATGCT 2107
                 |||   |||||||||:::|||      ||||||          ||||||
Db        155 IleGlnCysAlaSerLeuIleAlaTrpMetPhePheSerAspThrProArg--------- 171

Qy       2108 AAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGA 2167
                                              ::::::|||   :::   :::
Db        172 ---------------------------------LeuSerPheIleIleIleLeuGlnLys 180

Qy       2168 CGTTCATCAAGGAAACCGGAGAAACTTAAAACGCTCTTCTGCTACTTTAGAAGAGTCACA 2227
                 :::            ||||||||||||   ||||||   ||||||   :::::::::
Db        181 ThrThrCysValAlaValGluLysLeuLysPheLeuPheThrTyrPheAspLysMetSer 200

Qy       2228 GAGAAAAAACCTACTGGGTTGGTGACATTTACA------------------AGACAGAGT 2269
                 |||   |||   |||:::|||                           :::|||
Db        201 IleAspProProIleGlyAlaValSerPheArgLysMetArgIleThrHisLysGlnTyr 220

Qy       2270 CTTGAAGATTTTCCAGAATGGGAA---AGATGTGAAAAACCCTTGACACGATTGCATGTC 2326
              ||||||:::         |||:::   |||               |||   :::   |||
Db        221 LeuGluAsn---------TrpLysLeuArgGluThrAsnLeuLeuProAspValGlnVal 237

Qy       2327 ACTTACGAAGGTACCATAGAAGAAAATGGCCAAGGCATGCTACAGGTGGATTTTGCAAAT 2386
                 :::   :::|||||||||                       |||:::||||||||||||
Db        238 PheAspLysMetSerIleGluGluThrAlaLeu---CysThrGlnIleAspPheAlaAsn 256

Qy       2387 CGTTTTGTTGGAGGTGGTGTAACCAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTTA 2446
              :::   :::|||||||||||||            |||   ||||||||||||||||||||:::
Db        257 LysArgLeuGlyGlyGlyValLeuLysGlyGlyAlaValGlnGluGluIleArgPheMet 276

Qy       2447 ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGGATCACAATGAATGT 2506
              :::   ||||||::::::::::::         |||      :::|||            |||
Db        277 MetCysProGluMetMetValAlaIleLeuLeuAsnAspValThrGlnAspLeuGluAla 296

Qy       2507 CTAATTATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566
              :::   |||   |||         :::|||   ||||||||||||:::   |||   :::
Db        297 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrSerAsnThrLeuLys 316

Qy       2567 TGGTCCCGG------------------AGCCACGAAGATGGGAGTGAAAGGGACGACTGC 2608
              |||::::::                  :::      :::   |||   ||||||
Db        317 TrpAlaLysIleThrProLysHisSerAlaGlnAsnAsnAsnSerPheArgAspGlnPhe 336

Qy       2609 GAGCGGCGCTGCACTGAGATCGTTGCCATCGATGCTCTTCAC---------------TTC 2653
                 |||      ||||||   ||||||||||||||||:::
Db        337 GlyArgLeuGlnThrGluThrValAlaIleAspAlaValArgAsnAlaGlyThrProLeu 356

Qy       2654 AGACGCTACCTCGATCAGTTTGTGCCTGAGAAAATGAGACGCGAGCTGAACAAGGCTTAC 2713
```

```
                       |||:::|||         ||||||:::   ||||||:::   ||||||
Db        357 GluCysLeuLeuAsnGlnLeuThrThrGluLysLeuThrArgGluValArgLysAlaAla 376

Qy       2714 TGTGGATTTCTCCGTCCTGGAGTTTCTTCAGAGAATCTTTCTGCAGTGGCCACAGGAAAC 2773
                 |||||||||      |||               :::      |||   :::|||
Db        377 IleGlyPheLeuSerAlaGlyAspGlyPheSerLysIle---ProValValSerGlyTrp 395

Qy       2774 TGGGGCTGTGGTGCCTTTGGGGGTGATGCCAGGTTAAAAGCCTTAATACAGATATTGGCA 2833
              ||||||||||||||||||   |||:::      ||||||   |||||||||::::::|||
Db        396 TrpGlyCysGlyAlaPheArgGlyAsnLysProLeuLysPheLeuIleGlnValIleAla 415

Qy       2834 GCTGCTGCAGCTGAGCGAGATGTGGTTTATTTCACCTTTGGGGACTCAGAATTGATGAGA 2893
                       ::::::|||   :::   :::   ||||||||||||:::||||||   :::
Db        416 CysGlyIleSerAspArgProLeuGlnPheCysThrPheGlyAspThrGluLeuAlaLys 435

Qy       2894 GACATTTACAGCATGCACATTTTCCTTACTGAAAGGAAACTCACTGTTGGAGATGTGTAT 2953
                          |||                        :::      |||   ::::::
Db        436 LysCysGluGluMetMetThrLeuPheArgAsnAsnAsnValArgThrGlyGlnLeuPhe 455

Qy       2954 AAGCTGTTG 2962
                 ::::::
Db        456 LeuIleIle 458
```

Search completed: May 26, 2004, 17:08:53
Job time : 291.952 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May 26, 2004, 17:19:25 ; Search time 111.821 Seconds
                                          (without alignments)
                                          2466.140 Million cell updates/sec

Title:         US-09-302-812-4
Perfect score: 5190
Sequence:      1 MNAGPGCEPCTKATRWGAAT..........YHAVESCAETADHSGQRTGT 976

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:     1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
               1:  geneseqp1980s:*
               2:  geneseqp1990s:*
               3:  geneseqp2000s:*
               4:  geneseqp2001s:*
               5:  geneseqp2002s:*
               6:  geneseqp2003as:*
               7:  geneseqp2003bs:*
               8:  geneseqp2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	100.0	976	5	AAE25630	Aae25630 Human pol
2	5190	100.0	976	5	AAU76021	Aau76021 Human pol
3	5190	100.0	976	5	AAU76012	Aau76012 Human pol
4	5190	100.0	976	6	ABG72279	Abg72279 Human pol
5	4635.5	89.3	977	5	AAE25629	Aae25629 Bovine po
6	4635.5	89.3	977	5	AAU76020	Aau76020 Bovine po
7	4635.5	89.3	977	5	AAU75799	Aau75799 Bovine po
8	4635.5	89.3	977	6	ABG72278	Abg72278 Bovine po
9	4395	84.7	968	5	AAE25631	Aae25631 Murine po
10	4395	84.7	968	5	AAU76022	Aau76022 Mouse pol
11	4395	84.7	968	5	AAU76013	Aau76013 Mouse pol
12	4395	84.7	968	6	ABG72280	Abg72280 Mmurine p
13	1078	20.8	768	4	ABB59491	Abb59491 Drosophil
14	1078	20.8	768	5	AAE25632	Aae25632 Fruit fly
15	1078	20.8	768	5	AAU76023	Aau76023 Fruit fly
16	1078	20.8	768	5	AAU76014	Aau76014 Fruit fly
17	1078	20.8	768	6	ABG72281	Abg72281 Fruit fly
18	463	8.9	726	5	AAE25633	Aae25633 Poly aden
19	463	8.9	726	5	AAU76024	Aau76024 Worm poly
20	463	8.9	726	5	AAU76015	Aau76015 Worm poly
21	463	8.9	726	6	ABG72282	Abg72282 C. elegan
22	344.5	6.6	819	4	ABG20721	Abg20721 Novel hum
23	210	4.0	100	4	ABG11103	Abg11103 Novel hum
24	178	3.4	33	5	AAE25652	Aae25652 Bovine po
25	174	3.4	4873	6	ABO14747	Abo14747 Novel hum
26	173.5	3.3	748	4	AAM58340	Aam58340 Human bra
27	170	3.3	335	4	ABG09631	Abg09631 Novel hum
28	170	3.3	2703	4	ABB60074	Abb60074 Drosophil
29	167	3.2	31	5	AAE25634	Aae25634 Bovine po
30	167	3.2	31	5	AAE25651	Aae25651 Bovine po
31	167	3.2	31	5	AAU76025	Aau76025 Bovine po
32	167	3.2	31	5	AAU76016	Aau76016 Bovine po
33	167	3.2	31	6	ABG72283	Abg72283 Oligopept
34	167	3.2	1274	6	ABJ18980	Abj18980 Pathogen
35	167	3.2	1274	6	ABM71209	Abm71209 Staphyloc
36	167	3.2	1803	4	ABB65391	Abb65391 Drosophil
37	166.5	3.2	1163	4	AAU28028	Aau28028 Novel hum
38	166.5	3.2	2194	4	AAM40114	Aam40114 Human pol
39	165.5	3.2	1274	6	ABU15913	Abu15913 Protein e
40	164.5	3.2	1111	4	AAU37362	Aau37362 Staphyloc
41	162	3.1	768	4	AAM78686	Aam78686 Human pro
42	160	3.1	1212	4	ABG04004	Abg04004 Novel hum
43	160	3.1	1213	4	ABG23201	Abg23201 Novel hum
44	160	3.1	1213	4	ABG18490	Abg18490 Novel hum
45	160	3.1	1241	4	ABG14773	Abg14773 Novel hum

ALIGNMENTS

```
RESULT 1
AAE25630
ID   AAE25630 standard; protein; 976 AA.
XX
AC   AAE25630;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Human poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Homo sapiens.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42082.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 55-60; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
```

```
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is human PARG
XX
SQ   Sequence 976 AA;

  Query Match             100.0%;  Score 5190;  DB 5;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
                ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976


RESULT 2
AAU76021
ID   AAU76021 standard; protein; 976 AA.
XX
AC   AAU76021;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Homo sapiens.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14932.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 55-60; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the human PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 976 AA;

  Query Match             100.0%;  Score 5190;  DB 5;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
              ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976


RESULT 3
AAU76012
ID   AAU76012 standard; protein; 976 AA.
XX
AC   AAU76012;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Homo sapiens.
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14494.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 55-60; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the human PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 976 AA;

  Query Match             100.0%;  Score 5190;  DB 5;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
              ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976

RESULT 4
ABG72279
ID   ABG72279 standard; protein; 976 AA.
XX
AC   ABG72279;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14478.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents human PARG enzyme
XX
SQ   Sequence 976 AA;

  Query Match             100.0%;  Score 5190;  DB 6;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
              ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976

RESULT 5
AAE25629
ID   AAE25629 standard; protein; 977 AA.
XX
AC   AAE25629;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).
XX
KW   Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Bos taurus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42081.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 47-45; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is bPARG
XX
SQ   Sequence 977 AA;

  Query Match             89.3%;  Score 4635.5;  DB 5;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:||||||||||  || || || |||||||||| ||||||| ||| |||||||||  |  
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:||||||||||||||||||||||||||||| ||||| |||||||| |||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
              |||:||| |||||  |||  || |||| |||||||||||||| |:|:||||||  |||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              |||||||:|:|   ||||||:|||:|||||||| ||||||||| |||  |||:||||| |
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              :||| | || |||:| ||| :||||| |||||||| || |:| |||| ||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              |||||||||||||||||||||||||||||||| ||:|||||||||| |:|| | ||| | 
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              ||||:|||||||||||| ||||:||| ||: ||||||||||||||:|||||||| ||:||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              |||:|||| | ||||||||||||:||||||||||||||||||||||||||||| ||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||||||||| ||||||||||:|||||||||||||||||||||||||||| ||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              ||||||: ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600
```

```
Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
              |||||||| ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
              ||||||||||||||||||||||:||||| ||||| :|| |||||||||||||||||||||
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
              |:||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
              |||||||||||||||| ||||||||||:||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
              |||| :| :  |||||
Db        961 VESCTQTTNQPGQRTG 976
```

RESULT 6
AAU76020
ID   AAU76020 standard; protein; 977 AA.
XX
AC   AAU76020;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Bos taurus.
XX
FH   Key   Location/Qualifiers
FT   Region   601. .617
FT   /note= "Represents PARG oligopeptide #1"
FT   Region   761. .770
FT   /note= "Represents PARG oligopeptide #2"
FT   Region   771. .801
FT   /note= "Represents PARG oligopeptide #3"
FT   Region   849. .880
FT   /note= "Represents PARG oligopeptide #4"
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14931.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Claim 2; Col 47-52; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the bovine PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 977 AA;

```
  Query Match             89.3%;  Score 4635.5;  DB 5;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:||||||||||  || || || |||||||||| ||||||| ||| |||||||||  |  
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:||||||||||||||||||||||||||||| ||||| |||||||| |||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
              |||:||| |||||  |||  || |||| |||||||||||||| |:|:||||||  |||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              |||||||:|:|   ||||||:|||:|||||||| ||||||||| |||  |||:||||| |
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              :||| | || |||:| |||  ||||| |||||||| || |:| |||| ||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              |||||||||||||||||||||||||||||||| ||:|||||||||| |:|| | ||| | 
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              ||||:|||||||||||| ||||:||| ||: ||||||||||||||:|||||||| ||:||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              |||:|||| | ||||||||||||:||||||||||||||||||||||||||||| ||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||||||||| ||||||||||:|||||||||||||||||||||||||||| ||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              |||||: |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660
```

```
Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976

RESULT 7
AAU75799
ID   AAU75799 standard; protein; 977 AA.
XX
AC   AAU75799;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW   adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW   inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW   neurodegeneration; Huntington's disease; Parkinson's disease;
KW   Alzheimer's disease; neurotoxicity.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Region          601..617
FT                   /note= "Represents PARG oligopeptide #1"
FT   Region          761..770
FT                   /note= "Represents PARG oligopeptide #2"
FT   Region          771..801
FT                   /note= "Represents PARG oligopeptide #3"
FT   Region          849..880
FT                   /note= "Represents PARG oligopeptide #4"
XX
PN   US6333148-B1.
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14493.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 45-52; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the bovine PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 977 AA;

  Query Match             89.3%;  Score 4635.5;  DB 5;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
```

```
Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976
```

```
RESULT 8
ABG72278
ID   ABG72278 standard; protein; 977 AA.
XX
AC   ABG72278;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW   neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;
KW   anticonvulsant; cerebroprotective; enzyme.
XX
OS   Bos taurus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14477.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents bovine PARG enzyme
XX
SQ   Sequence 977 AA;

  Query Match            89.3%;  Score 4635.5;  DB 6;  Length 977;
  Best Local Similarity  89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900
```

```
Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
              |||||||||||||||| ||||||||||:||||||||||||||||||||||||||||||||
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
              ||||  :|  :  |||||
Db        961 VESCTQTTNQPGQRTG 976


RESULT 9
AAE25631
ID   AAE25631 standard; protein; 968 AA.
XX
AC   AAE25631;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy.
XX
OS   Mus musculus.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42083.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Claim 3; Col 63-68; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is murine PARG
XX
SQ   Sequence 968 AA;

  Query Match             84.7%;  Score 4395;  DB 5;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:|||  ||||||  |||||  |::|  |||:||||  ||||||||||  ||||||||||||
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:|| |||:||| |||||||:|||||| ||||||    ||||||||:||||||||||||
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
               | |||| || ||:|||| ||||:||||| |||::|||||||| |||| |||    | :|
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              :| ||||| :||  |||||| ||:|||||||| :|||| |  |   |  ||   |: |||
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              |    :||  |||:|:||:|:|||||||:||:||| ||||| ||| ||:||||||||
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              |||||||||:||||||||||||||| ||||:|  ||||||| |   |||:|| : |||| |
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              || ||||||||| ||:  | :||||| || ||||||||:|||||:||||||| ||: |
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              :|||||| | :|||||||||||:||:| |:|||||||||||:||||| || |||||:
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
              ||||||| ||:|||||||||||||||| |||||||||||||||||:||||||||||||
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
              |||||||||||||||||||:||||||||:||  :|||||||||||||||||||||||||
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
              |:|||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
              |||||||||||||||| ||||||| || |||||||||||||||||||||||||||||||
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
              ||| |||  |   ||: ||
Db        952 VESSAETTDMPGQKAGT 968


RESULT 10
AAU76022
ID   AAU76022 standard; protein; 968 AA.
XX
AC   AAU76022;
```

XX
DT 08-MAY-2002 (first entry)
XX
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease.
XX
OS Mus musculus.
XX
PN US6337202-B1.
XX
PD 08-JAN-2002.
XX
PF 23-FEB-2000; 2000US-00511477.
XX
PR 01-MAY-1998; 98US-0083768P.
PR 30-APR-1999; 99US-00302812.
XX
PA (KENT) UNIV KENTUCKY RES FOUND.
XX
PI Jacobson MK, Jacobson EL, Ame J, Lin W;
XX
DR WPI; 2002-163240/21.
DR N-PSDB; ABK14933.
XX
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS Claim 2; Col 63-70; 81pp; English.
XX
CC The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC (PARG) protein which catalyses release of ADP-ribose from an ADP
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC invention is useful for generating antibodies and can be inhibited or
CC activated for diagnosing and treating neoplastic disorders such as
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC and related conditions. PARG levels may be enhanced to suppress DNA
CC repair and increase the cell's susceptibility to chemotherapy drugs.
CC Antagonists of PARG are administered to treat or prevent neoplastic
CC disorders. The present amino acid sequence represents the mouse PARG
CC protein of the invention. This protein is one of several PARG proteins
CC (AAU76020-AAU76024) of the invention
XX
SQ Sequence 968 AA;

```
  Query Match             84.7%;  Score 4395;  DB 5;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:|||| |||||| ||||| |::| |||:|||| ||||||||||| ||||||||||||| 
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:|| |||:||| ||||||||:|||||| ||||||   |||||||||:||||||||||||
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
               | ||||  || ||:|||| ||||:||||| |||::||||||||| |||| |||   | :|
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              :| ||||| :||  ||||||| ||:|||||||| :|||| | |  |  ||   |: ||| 
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              |    :||  |||:|:||:|:|||||||:||:||| ||||||  ||| ||:|||||||||
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              ||||||||||:|||||||||||||| ||||:|  ||||||| |  |||:|| : |||| |
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              || ||||||||| ||:  | :|||||  || ||||||||:||||||:||||||| ||: |
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              :||||||| | :|||||||||||:||:| |:||||||||||||:|||||| || |||||:
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
              |||||||| ||:||||||||||||||||| |||||||||||||||||:||||||||||||
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
              ||||||||||||||||||||||:||||||||:|| :||||||||||||||||||||||||
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
              |:|||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
              |||||||||||||||| ||||||| || ||||||||||||||||||||||||||||||||
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
              ||| ||| |  ||: ||
Db        952 VESSAETTDMPGQKAGT 968
```

RESULT 11
AAU76013
ID AAU76013 standard; protein; 968 AA.
XX
AC AAU76013;
XX
DT 08-MAY-2002 (first entry)
XX
DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW neurodegeneration; Huntington's disease; Parkinson's disease;
KW Alzheimer's disease; neurotoxicity.
XX
OS Mus musculus.
XX
PN US6333148-B1.

```
XX
PD   25-DEC-2001.
XX
PF   30-APR-1999;   99US-00302812.
XX
PR   01-MAY-1998;   98US-0083768P.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-153820/20.
DR   N-PSDB; ABK14495.
XX
PT   Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT   useful potentially for treating diseases associated with DNA damage, e.g.
PT   cancer.
XX
PS   Claim 3; Col 63-68; 80pp; English.
XX
CC   The present invention relates to a new method for screening compounds for
CC   ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC   diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC   inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC   are used to treat or prevent any condition associated with DNA damage
CC   e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC   vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC   Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC   Compounds identified by the new method are more effective than known
CC   inhibitors and have fewer side effects. The present amino acid sequence
CC   represents the mouse PARG protein of the invention. This protein is one
CC   of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC   invention
XX
SQ   Sequence 968 AA;

  Query Match             84.7%;  Score 4395;  DB 5;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
Db        952 VESSAETTDMPGQKAGT 968

RESULT 12
ABG72280
ID   ABG72280 standard; protein; 968 AA.
XX
AC   ABG72280;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX
KW   Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW   cellular response; DNA damage; neoplastic disorder inducing agent;
KW   oxidative stress; neoplastic disorder; myocardial infarction;
KW   vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; inborn genetic error;
KW   reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;
KW   cytostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;
KW   vasotropic; anticonvulsant; cerebroprotective; enzyme.
XX
OS   Mus musculus.
XX
PN   US2002132328-A1.
XX
PD   19-SEP-2002.
XX
PF   09-OCT-2001; 2001US-00973451.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (JACO/) JACOBSON M K.
PA   (JACO/) JACOBSON E L.
PA   (AMEJ/) AME J.
PA   (LINW/) LIN W.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
```

```
DR   WPI; 2003-155895/15.
DR   N-PSDB; ABX14479.
XX
PT   New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT   glycohydrolase activity, for preventing, treating, or ameliorating a
PT   disease condition, e.g. neoplastic disorder, myocardial infarction or
PT   vascular stroke.
XX
PS   Claim 28; Fig 16; 86pp; English.
XX
CC   The present invention relates to the isolation of poly(ADP-ribose)
CC   glycohydrolase (PARG) from several species, and the polynucleotide
CC   sequences encoding them. Methods for inhibiting PARG expression or
CC   overexpressing PARG are also disclosed. PARG is involved in the cellular
CC   response to DNA damage, and is associated with the body's response to
CC   neoplastic disorder inducing agents and oxidative stress. The
CC   polynucleotide sequences encoding PARG and PARG modulators are useful for
CC   preventing, treating, or ameliorating diseases such as neoplastic
CC   disorders, myocardial infarction, vascular stroke, neurodegenerative
CC   disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC   and neurotoxicity. The polynucleotide sequences are also useful in gene
CC   therapy. The methods are useful for identifying an agent that modulates
CC   PARG activity, identifying a mutant PARG allele in an individual, or
CC   screening candidate molecules for PARG modulating activity. The present
CC   sequence represents murine PARG enzyme
XX
SQ   Sequence 968 AA;

  Query Match             84.7%;  Score 4395;  DB 6;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
Db        952 VESSAETTDMPGQKAGT 968


RESULT 13
ABB59491
ID   ABB59491 standard; protein; 768 AA.
XX
AC   ABB59491;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5265.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03594.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
```

```
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1078;  DB 4;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7e-84;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
              |: |: |      : ||: :                | :|  |  |  |  |  :| :   |  |     |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
              | :       | |   :: :       |:|: :  |   ||::||::||: :: || |: :|  |
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
                  |||:|: |||        : |: ||: ||   |   :|  | ||      |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
               :|| ::::|| ||::    |:||||    | |:::||:||: |||||| |||||||   |  |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
              ||||::||||||||::              |||| :  |||||  ||:         |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
              |   :| :    ||       |||  |||||: | |:||||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
              | |||:: :|||| |     | |:: | |:|| ||||| :: || : || : ||     || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
              |||||||| :     |:  :  | |||||| ||:               ||           |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
              ||||: ||||:|::    |       |  :  |:||  |   | |  :                ||    :: :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
               |:   :  |:  |              ||| ||
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589

RESULT 14
AAE25632
ID   AAE25632 standard; protein; 768 AA.
XX
AC   AAE25632;
XX
DT   04-NOV-2002  (first entry)
XX
DE   Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).
XX
KW   Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;
KW   ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;
KW   cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;
KW   Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;
KW   cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;
KW   antisense therapy; fruit fly.
XX
OS   Drosophila melanogaster.
XX
PN   US6395543-B1.
XX
PD   28-MAY-2002.
XX
PF   23-FEB-2000; 2000US-00511507.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-535641/57.
DR   N-PSDB; AAD42084.
XX
PT   New nucleic acid molecule encoding bovine poly adenosine diphosphate-
PT   ribose glycohydrolase involved in cellular response to DNA damage,
PT   inhibition of which is useful for treating neoplastic disorders and
PT   neurodegenerative diseases.
XX
PS   Example 13; Col 71-76; 77pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule which encodes
CC   a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)
CC   which catalyses release of ADP-ribose from an ADP ribose polymer. The
CC   invention is useful as probes and primer molecules that can used in
CC   hybridisation assays and polymerase chain reaction (PCR) amplification.
CC   The knowledge of the nucleotide sequence of the PARG gene permits the
CC   preparation of antisense therapeutics containing sequences complementary
CC   to the mRNA of PARG gene. The antisense therapeutic are useful to treat
CC   neoplastic disorders and conditions caused by genotoxic oxidative stress
CC   e.g., cardiac disorders, neuronal disorders, reperfusion injury,
CC   neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's
CC   disease. The invention is useful in gene therapy and antisense therapy.
CC   The present sequence is fruit fly PARG
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1078;  DB 5;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7e-84;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
              |: |: |      : ||: :                | :|  |  |  |  |  :| :   |  |     |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
              | :       | |   :: :       |:|: :  |   ||::||::||: :: || |: :|  |
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
                  |||:|: |||        : |: ||: ||   |   :|  | ||      |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
               :|| ::::|| ||::    |:||||    | |:::||:||: |||||| |||||||   |  |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
              ||||::||||||||::              |||| :  |||||  ||:         |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
              |   :| :    ||       |||  |||||: | |:||||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
              | |||:: :|||| |     | |:: | |:|| ||||| :: || : || : ||     || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
              |||||||| :     |:  :  | |||||| ||:               ||           |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
```

```
              ||||: ||||:|:: |    | : |:|||: |    | : | :         ||  ::  :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
               |:   :  |:  |          ||| ||
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589

RESULT 15
AAU76023
ID   AAU76023 standard; protein; 768 AA.
XX
AC   AAU76023;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Fruit fly poly(ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
KW   Fruit fly; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW   adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW   leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW   neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW   Huntington's disease; Parkinson's disease.
XX
OS   Drosophila melanogaster.
XX
PN   US6337202-B1.
XX
PD   08-JAN-2002.
XX
PF   23-FEB-2000; 2000US-00511477.
XX
PR   01-MAY-1998;   98US-0083768P.
PR   30-APR-1999;   99US-00302812.
XX
PA   (KENT ) UNIV KENTUCKY RES FOUND.
XX
PI   Jacobson MK,  Jacobson EL,  Ame J,  Lin W;
XX
DR   WPI; 2002-163240/21.
DR   N-PSDB; ABK14934.
XX
PT   Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
PT   which catalyses release of ADP-ribose from ADP ribose polymer, useful for
PT   treating neoplastic and neurological disorders, heart attack and stroke.
XX
PS   Example 13; Col 71-76; 81pp; English.
XX
CC   The present invention relates to a new poly(ADP-ribose) glycohydrolase
CC   (PARG) protein which catalyses release of ADP-ribose from an ADP
CC   (adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC   invention is useful for generating antibodies and can be inhibited or
CC   activated for diagnosing and treating neoplastic disorders such as
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC   ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC   disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC   and related conditions. PARG levels may be enhanced to suppress DNA
CC   repair and increase the cell's susceptibility to chemotherapy drugs.
CC   Antagonists of PARG are administered to treat or prevent neoplastic
CC   disorders. The present amino acid sequence represents the fruit fly PARG
CC   protein of the invention. This protein is one of several PARG proteins
CC   (AAU76020-AAU76024) of the invention
XX
SQ   Sequence 768 AA;

  Query Match             20.8%;  Score 1078;  DB 5;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7e-84;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
               |: |: |     : ||: :              | :|  |  |  |  |  :| :   | |      |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
               | :      | |  :: :       |:|: :  |  ||::||::||: :: || |: :| |
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
                   |||:|: |||       : |: ||: ||  |   :| | ||    |: |:|:| :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
                :|| ::::|| ||::   |:|||    | |:::||:||: |||||| ||||||   | |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
              ||||::||||||||::          |||| :    |||||   ||:        |||:||| |:|
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
              |   :| :     ||       |||  |||||: | |:||||||:::||||     | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
              | |||:: :||||  |    | |:: | |:|| ||||| :: || : || : ||    || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
              |||||||||  :     |:  : | ||||||| ||:          ||          |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
              ||||: ||||:|::   |    | : |:|||: |    | : | :          ||  ::  :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
               |:   :  |:  |          ||| ||
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589

Search completed: May 26, 2004, 18:40:01
Job time : 116.821 secs
```

```
                  GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 26, 2004, 18:38:05 ; Search time 31.2745 Seconds
                                          (without alignments)
                                          1611.120 Million cell updates/sec

Title:          US-09-302-812-4
Perfect score:  5190
Sequence:       1 MNAGPGCEPCTKATRWGAAT..........YHAVESCAETADHSGQRTGT 976

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	100.0	976	4	US-09-302-812-4	Sequence 4, Appli
2	5190	100.0	976	4	US-09-511-477-4	Sequence 4, Appli
3	5190	100.0	976	4	US-09-511-507-4	Sequence 4, Appli
4	4635.5	89.3	977	4	US-09-302-812-2	Sequence 2, Appli
5	4635.5	89.3	977	4	US-09-511-477-2	Sequence 2, Appli
6	4635.5	89.3	977	4	US-09-511-507-2	Sequence 2, Appli
7	4395	84.7	968	4	US-09-302-812-6	Sequence 6, Appli
8	4395	84.7	968	4	US-09-511-477-6	Sequence 6, Appli
9	4395	84.7	968	4	US-09-511-507-6	Sequence 6, Appli
10	1078	20.8	768	4	US-09-302-812-8	Sequence 8, Appli
11	1078	20.8	768	4	US-09-511-477-8	Sequence 8, Appli
12	1078	20.8	768	4	US-09-511-507-8	Sequence 8, Appli
13	463	8.9	726	4	US-09-302-812-10	Sequence 10, Appl
14	463	8.9	726	4	US-09-511-477-10	Sequence 10, Appl
15	463	8.9	726	4	US-09-511-507-10	Sequence 10, Appl
16	167	3.2	31	4	US-09-302-812-11	Sequence 11, Appl
17	167	3.2	31	4	US-09-511-477-11	Sequence 11, Appl
18	167	3.2	31	4	US-09-511-507-11	Sequence 11, Appl
19	154	3.0	1972	4	US-08-875-435B-3	Sequence 3, Appli
20	150.5	2.9	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
21	150.5	2.9	1972	4	US-08-875-435B-4	Sequence 4, Appli
22	150	2.9	3969	3	US-08-061-376-5	Sequence 5, Appli
23	149.5	2.9	1187	1	US-08-320-559-28	Sequence 28, Appl
24	149.5	2.9	1187	3	US-08-545-860D-28	Sequence 28, Appl
25	149.5	2.9	1187	5	PCT-US94-04496-28	Sequence 28, Appl
26	149.5	2.9	1210	1	US-08-320-559-26	Sequence 26, Appl
27	149.5	2.9	1210	3	US-08-545-860D-26	Sequence 26, Appl
28	149.5	2.9	1210	5	PCT-US94-04496-26	Sequence 26, Appl
29	149	2.9	26	4	US-09-302-812-20	Sequence 20, Appl
30	149	2.9	26	4	US-09-511-477-20	Sequence 20, Appl
31	149	2.9	26	4	US-09-511-507-20	Sequence 20, Appl
32	149	2.9	29	4	US-09-302-812-12	Sequence 12, Appl
33	149	2.9	29	4	US-09-511-477-12	Sequence 12, Appl
34	149	2.9	29	4	US-09-511-507-12	Sequence 12, Appl
35	149	2.9	427	4	US-09-134-001C-5143	Sequence 5143, Ap
36	147	2.8	754	4	US-09-976-594-375	Sequence 375, App
37	146.5	2.8	2125	4	US-09-919-172-29	Sequence 29, Appl
38	144	2.8	1596	4	US-08-978-277A-4	Sequence 4, Appli
39	142	2.7	2008	4	US-09-091-501B-8	Sequence 8, Appli
40	142	2.7	3433	4	US-09-091-501B-10	Sequence 10, Appl
41	141	2.7	1588	5	PCT-US93-07261-11	Sequence 11, Appl
42	141	2.7	1663	5	PCT-US93-07261-16	Sequence 16, Appl
43	139.5	2.7	1279	4	US-09-724-517-2	Sequence 2, Appli
44	139.5	2.7	1279	4	US-09-641-807A-2	Sequence 2, Appli
45	139.5	2.7	1279	4	US-09-723-096-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-302-812-4

  Query Match             100.0%;  Score 5190;  DB 4;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy          361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy          421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy          481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy          541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy          601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy          661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy          721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy          781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy          841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy          901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy          961 ESCAETADHSGQRTGT 976
                ||||||||||||||||
Db          961 ESCAETADHSGQRTGT 976
```

```
RESULT 2
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,477
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-477-4
```

```
  Query Match             100.0%;  Score 5190;  DB 4;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy           61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy          121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy          181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy          241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy          301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy          361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy          421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy          481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy          541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy          601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy          661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy          721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy          781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy          841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy          901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy          961 ESCAETADHSGQRTGT 976
                ||||||||||||||||
Db          961 ESCAETADHSGQRTGT 976
```

```
RESULT 3
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-511-507-4

  Query Match             100.0%;  Score 5190;  DB 4;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
              ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976

RESULT 4
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-302-812-2

  Query Match             89.3%;  Score 4635.5;  DB 4;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:||||||||||  || || || |||||||||| ||||||| ||| |||||||||  |  
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:||||||||||||||||||||||||||||| ||||| |||||||| |||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
              |||:||| |||||  |||  || |||| |||||||||||||| |:|:||||||  |||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              |||||||:|:|   ||||||:|||:||||||||| |||||||||| |||  |||:|||||| |
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              :||| | || |||:| ||| :||||| |||||||| || |:| |||| ||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300
```

```
Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976
```

RESULT 5
US-09-511-477-2
; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:

US-09-511-477-2

```
  Query Match             89.3%;  Score 4635.5;  DB 4;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976
```

```
RESULT 6
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 2
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Bos taurus
;   FEATURE:
US-09-511-507-2

  Query Match             89.3%;  Score 4635.5;  DB 4;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976


RESULT 7
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-302-812-6

  Query Match             84.7%;  Score 4395;  DB 4;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
```

```
Db   234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy   300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db   294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy   360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db   353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy   420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db   412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy   480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db   472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy   540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db   532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy   600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db   592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy   660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db   652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy   720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db   712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy   780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db   772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy   840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db   832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy   900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db   892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy   960 VESCAETADHSGQRTGT 976
Db   952 VESSAETTDMPGQKAGT 968
```

RESULT 8
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

Query Match 84.7%; Score 4395; DB 4; Length 968;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

```
Qy     1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db     1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy    60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db    60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy   120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db   117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy   180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db   177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy   240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db   234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy   300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db   294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy   360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db   353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy   420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db   412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy   480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db   472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy   540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db   532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy   600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db   592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy   660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db   652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy   720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db   712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy   780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db   772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy   840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db   832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy   900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db   892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy   960 VESCAETADHSGQRTGT 976
```

```
                 ||| ||| |  ||: ||
Db        952 VESSAETTDMPGQKAGT 968


RESULT 9
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
;  APPLICANT:  JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AME, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION:  GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE:  NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/511,507
;  CURRENT FILING DATE:  2000-02-23
;  PRIOR APPLICATION NUMBER:  09/302,812
;  PRIOR FILING DATE:  1999-04-30
;  NUMBER OF SEQ ID NOS:  38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-511-507-6

  Query Match             84.7%;  Score 4395;  DB 4;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:|||| |||||| ||||| |::| |||:|||| |||||||||| ||||||||||||
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:|| |||:||| |||||||:|||||| |||||   |||||||:||||||||||
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
               |||| || ||:|||| ||||:||||| ||::||||||||| |||| |||   :|
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              :| ||||| :||  |||||| ||:||||||| :|||| |  |  ||    |: |||
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              |    :||  |||:|:||:||||||:||:|||  ||||||  ||| ||:||||||||
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              |||||||||:||||||||||||| ||||:|  |||||| |  |||:|| : |||  |
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              || ||||||||| ||:  | :||||| || |||||||:||||||:||||||| ||:  |
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              :|||||| | :||||||||||||:||:| |:|||||||||||||:||||||| || |||||:
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
              ||||||| ||:||||||||||||||||||| |||||||||||||||||:|||||||||||
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
              ||||||||||||||||||||:|||||||||:|| :|||||||||||||||||||||||
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
              |:||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
              ||||||||||||||| ||||||| || ||||||||||||||||||||||||||||||||
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
              ||| ||| |  ||: ||
Db        952 VESSAETTDMPGQKAGT 968


RESULT 10
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT: JACOBSON, Elaine L.
;  APPLICANT: AM, Jean-Christophe
;  APPLICANT: LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
;  TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
;  TITLE OF INVENTION: THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/302,812B
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 60/083,768
;  EARLIER FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-302-812-8

  Query Match             20.8%;  Score 1078;  DB 4;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7.9e-93;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
              |: |: |     : ||: :              | :|  |  |  |  :|:   |  |
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
              | :    | |   :: :      |:|: :   |   ||::||::||: :: || |: :|  |
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
                 |||:|: |||       : |: ||: ||   |  :| | ||     |: |:|:|  :  ::
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
                :|| ::::|| ||::   |:||||    | |:::||:||: |||||  |||||||  | |
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270
```

```
Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589
```

RESULT 11
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-477-8

```
  Query Match             20.8%;  Score 1078;  DB 4;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7.9e-93;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
```

```
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589
```

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI\
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

```
  Query Match             20.8%;  Score 1078;  DB 4;  Length 768;
  Best Local Similarity   42.3%;  Pred. No. 7.9e-93;
  Matches  240;  Conservative   94;  Mismatches  184;  Indels   50;  Gaps   14;

Qy        416 RLPKAEDRRKEQWETKHQRTERKIPKYVPPHL--SPDKKWLGTPIEEM--RRMPRCGIRL 471
Db         45 RMSKSPDGGISEIETEEE----------PENLANSLDDSWRGVSMEAIHRNRQPFELENL 94

Qy        472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERT 531
Db         95 PPVTAGNLHRVMYQLPIRET--PPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGSTT 150

Qy        532 AGSRWELIQTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQ 591
Db        151 IDFRWEMIERALLQPIKTCEELQAAIISYNTTYRDQWHFRALHQLLDEELDESETRVFFE 210

Qy        592 SILPDMVKIALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNA-KMK 650
Db        211 DLLPRIIRLALRLPDLIQSPVPLLKHHKNASLSLSQQQISCLLANAFLCTFPRRNTLKRK 270

Qy        651 SEYSSYPDINFNRLFEGRSSRKPEKLKTLFCYFRRV--TEKK----PTGLVTFTRQS--L 702
Db        271 SEYSTFPDINFNRLYQSTGPAVLEKLKCIMHYFRRVCPTERDASNVPTGVVTFVRRSGLP 330

Qy        703 EDFPEWERCEKPL--TRLHVTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFL 760
```

```
                 |  :| :   ||     |||  |||||: | |:|||||||:::||||    | |||||||:
Db        331 EHLIDWSQSAAPLGDVPLHVDAEGTIEDEGIGLLQVDFANKYLGGGVLGHGCVQEEIRFV 390

Qy        761 INPELIISRLFTEVLDHNECLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTE 820
              | |||::  :||||  |     | |::  | :||  |||||  :: ||  :  || :  ||      || |
Db        391 ICPELLVGKLFTECLRPFEALVMLGAERYSNYTGYAGSFEWSGNFEDSTPRDSSGRRQTA 450

Qy        821 IVAIDALHFRRYLDQFVPEKMRRELNKAYCGFLR------PGVSSENLSAVATGNWGCGA 874
              ||||||||| :       |:  :  | ||||||| ||:           ||          |||||||||
Db        451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHWMVTPPPG--------VATGNWGCGA 502

Qy        875 FGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLR 934
              ||||: ||||:|::  |    |  :  |:|||: |    | : | :         ||  :: :||
Db        503 FGGDSYLKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMWLLFRNDGTTVQQLWS-ILR 561

Qy        935 YYNEECRNCSTPGP------DIKLYPFI 956
               |:  :  |: |          ||| ||
Db        562 SYSRLIKEKSSKEPRENKASKKKLYDFI 589
```

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-10

```
  Query Match             8.9%;  Score 463;  DB 4;  Length 726;
  Best Local Similarity   23.9%;  Pred. No. 1.5e-34;
  Matches  207;  Conservative  123;  Mismatches  278;  Indels  258;  Gaps   37;

Qy        145 YLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPEQFSNANIDRSPQNDDHSDTDSEEN 204
               : :|  |    |    |  ||    :               ||:  : ::  | |: |  |  || ||
Db         27 FAHQVPTMKRRKLTEHGNTTESKED---------PEEPKSRDVFVSSQSSDESQEDSAEN 77

Qy        205 -------RDNQQFLT-TVKLANAKQ----TTEDEHAREAKSHQKCSKSCHPGEDCASCQQ 252
                     :| : || |:|::| :          |    ||   :   :|      ||   | |:
Db         78 PEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLD--NH----KSTEPMEE------ 125

Qy        253 DEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEPESPMDVDNSKNS 312
                 || ||  :           ||     |  |::|: :|:        ||         |  :   :
Db        126 ---DVNNKSNI------DVAINSDEDDELVLEEN--------NKEMR-----DGEQVQQL 163

Qy        313 CQDSEADEE---TSPGFDEQE------DGSSSQTANKPSRFQARDADIEF-------RKR 356
               ||  ||::        ||   :           |  :||  |      :  :||   |              |
Db        164 SQDLFADDQELIEYPGIMKDTTTQLDITDSEVETAQKMEMIEETEADSTFVGEDSKATKT 223

Qy        357 YSTKG---------------GEVRLH--------FQFEGGESRTGMNDLNAKLPG----- 388
                |                |  |::          |   |  :  |    |||   |
Db        224 VRTSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTEG-NLTLQPDLNKVDPDRNYRY 282

Qy        389 ----NISSLNVECRNSKQHGKKDSKITDHLMRLPKAEDRRKEQWETKHQRTE------RK 438
                   |  :   :  |    ::| |           : ||    :|  ::::: :| :         ||
Db        283 CTIPNFPASQGKLREDNRYGPK--------IVLP----QRWREFDSRGRRRDSYFYFKRK 330

Qy        439 IPKYVPPHLSPDKKWLGTPIEEMRRM-PRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKP 497
               :  |:  : :            :  |    |     :||  |
Db        331 LDGYLKCYKTTGYFMFVGLLHNMWEFDPDITYKLPALE--------------------- 368

Qy        498 FPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDA 556
                 :||::                                  ||:  :  :| ||    :
Db        369 --MYYKEM----------------------------SELVGREEVLEKFARVARIA-- 394

Qy        557 ILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLK 616
                                    |  |:   :  :|: :: |:
Db        395 ---------------------KTAEDILPERIYR-LVGDV------------------ 412

Qy        617 QKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKL 676
                   | |:|  :|  |:|:|  ||            |:    |  :|          | |:       |   |||
Db        413 ----ESATLSHKQCAALVARMFF-------ARPDSPFS------FCRILSSDKSICVEKL 455

Qy        677 KTLFCYFRRVTEKKPTGLVTF--TRQSLEDF-PEWERCEKPLTRLHVT--YEGTIEENGQ 731
               | || ||  :::     | | |:|   |:    :  |  ||:  :| |   |           ::  : |:
Db        456 KFLFTYFDKMSMDPPDGAVSFRLTKMDKDTFNEEWK--DKKLRSLPEVEFFDEMLIEDTA 513

Qy        732 GMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTEQYSE 791
                  |||||   :||||   :  | ||||||||:  ||:::   |    |  :      |   :  |    :|
Db        514 LCTQVDFANEHLGGGVLNHGSVQEEIRFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSS 573

Qy        792 YTGYAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 840
              ||||  | :|            :  | : ::   ||     |     | :||||: |:               :|
Db        574 YTGYGHTLKWAELQPNHSRQNTNEFRDRFGRLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
               : ||: ||   ||:   |        |:   :  || ||||||  ||    ||  :||::||    |:| : :
Db        634 IIREMKKASIGFMSQGPKFTNI-PIVTGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFC 692

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVG 926
               :||: ||          :     : :: :|:|
Db        693 SFGEPELAAKCKKIIERMKQKDVTLG 718
```

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-10

```
  Query Match             8.9%;  Score 463;  DB 4;  Length 726;
  Best Local Similarity   23.9%;  Pred. No. 1.5e-34;
  Matches  207;  Conservative  123;  Mismatches  278;  Indels  258;  Gaps   37;

Qy        145 YLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPEQFSNANIDRSPQNDDHSDTDSEEN 204
               : :|  |    |    |  ||    :               ||:  : ::  | |: |  |  || ||
Db         27 FAHQVPTMKRRKLTEHGNTTESKED---------PEEPKSRDVFVSSQSSDESQEDSAEN 77

Qy        205 -------RDNQQFLT-TVKLANAKQ----TTEDEHAREAKSHQKCSKSCHPGEDCASCQQ 252
                     :| : || |:|::| :          |    ||   :   :|      ||   | |:
Db         78 PEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLD--NH----KSTEPMEE------ 125
```

```
Qy        253 DEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEPESPMDVDNSKNS 312
                 ||  || :         ||   |   |::|: :|:              ||         | :  :
Db        126 ---DVNNKSNI------DVAINSDEDDELVLEEN--------NKEMR-----DGEQVQQL 163

Qy        313 CQDSEADEE---TSPGFDEQE------DGSSSQTANKPSRFQARDADIEF-------RKR 356
               ||  ||::       ||    :           |   :||  |       :  :||  |          |
Db        164 SQDLFADDQELIEYPGIMKDTTTQLDITDSEVETAQKMEMIEETEADSTFVGEDSKATKT 223

Qy        357 YSTKG---------------GEVRLH--------FQFEGGESRTGMNDLNAKLPG----- 388
                |                    |::                  |   |  :  |     |||    |
Db        224 VRTSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTEG-NLTLQPDLNKVDPDRNYRY 282

Qy        389 ----NISSLNVECRNSKQHGKKDSKITDHLMRLPKAEDRRKEQWETKHQRTE------RK 438
                   |   :    :   |     ::|  |            :  ||       :|   ::::: :|  :             ||
Db        283 CTIPNFPASQGKLREDNRYGPK--------IVLP----QRWREFDSRGRRRDSYFYFKRK 330

Qy        439 IPKYVPPHLSPDKKWLGTPIEEMRRM-PRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKP 497
               :   |:   :   :           :     |        |           :||  |
Db        331 LDGYLKCYKTTGYFMFVGLLHNMWEFDPDITYKLPALE---------------------- 368

Qy        498 FPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDA 556
                  :||::                                   ||:  :    :|  ||  |      :
Db        369 --MYYKEM-----------------------------SELVGREEVLEKFARVARIA-- 394

Qy        557 ILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLK 616
                                                |   |:    :  :|: :: |:
Db        395 ----------------------KTAEDILPERIYR-LVGDV------------------ 412

Qy        617 QKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKL 676
                     | |:|  :|  |:|:|   ||             |:     |  :|         | |:            |||
Db        413 ----ESATLSHKQCAALVARMFF-------ARPDSPFS------FCRILSSDKSICVEKL 455

Qy        677 KTLFCYFRRVTEKKPTGLVTF--TRQSLEDF-PEWERCEKPLTRLHVT--YEGTIEENGQ 731
               | ||  ||  :::     |  | |:|    |:      :   |     ||:    :|  |     |              ::    :  |:
Db        456 KFLFTYFDKMSMDPPDGAVSFRLTKMDKDTFNEEWK--DKKLRSLPEVEFFDEMLIEDTA 513

Qy        732 GMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTEQYSE 791
                      |||||    :||||   :   |  |||||||:  ||:::      |     |   :      |  :  |     |      :|
Db        514 LCTQVDFANEHLGGGVLNHGSVQEEIRFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSS 573

Qy        792 YTGYAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 840
               ||||    |  :|              :   |   :  :  :|   ||        |    :||||:  |:                :|
Db        574 YTGYGHTLKWAELQPNHSRQNTNEFRDRFGRLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
               :  ||:  ||    ||:    |         |:    :  ||   ||||||  ||     ||   :||::||     |:|    :    :
Db        634 IIREMKKASIGFMSQGPKFTNI-PIVTGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFC 692

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVG 926
               :||:  ||          :       :  ::  :|:|
Db        693 SFGEPELAAKCKKIIERMKQKDVTLG 718
```

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-507-10

```
  Query Match             8.9%;  Score 463;  DB 4;  Length 726;
  Best Local Similarity   23.9%;  Pred. No. 1.5e-34;
  Matches  207;  Conservative  123;  Mismatches  278;  Indels  258;  Gaps   37;

Qy        145 YLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPEQFSNANIDRSPQNDDHSDTDSEEN 204
               : :|            |      |  ||     :                  ||:   :  ::   |  |:  |   |  ||  ||
Db         27 FAHQVPTMKRRKLTEHGNTTESKED---------PEEPKSRDVFVSSQSSDESQEDSAEN 77

Qy        205 -------RDNQQFLT-TVKLANAKQ----TTEDEHAREAKSHQKCSKSCHPGEDCASCQQ 252
                          :|  :  ||  |:|::|  :           |      ||    :    :|        ||    |  |:
Db         78 PEIAKEVSENCENLTETLKISNIESLDNVTERSEHTLD--NH----KSTEPMEE------ 125

Qy        253 DEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEPESPMDVDNSKNS 312
                 ||  || :         ||   |   |::|: :|:              ||         | :  :
Db        126 ---DVNNKSNI------DVAINSDEDDELVLEEN--------NKEMR-----DGEQVQQL 163

Qy        313 CQDSEADEE---TSPGFDEQE------DGSSSQTANKPSRFQARDADIEF-------RKR 356
               ||  ||::       ||    :           |   :||  |       :  :||  |          |
Db        164 SQDLFADDQELIEYPGIMKDTTTQLDITDSEVETAQKMEMIEETEADSTFVGEDSKATKT 223

Qy        357 YSTKG---------------GEVRLH--------FQFEGGESRTGMNDLNAKLPG----- 388
                |                    |::                  |   |  :  |     |||    |
Db        224 VRTSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTEG-NLTLQPDLNKVDPDRNYRY 282

Qy        389 ----NISSLNVECRNSKQHGKKDSKITDHLMRLPKAEDRRKEQWETKHQRTE------RK 438
                   |   :    :   |     ::|  |            :  ||       :|   ::::: :|  :             ||
Db        283 CTIPNFPASQGKLREDNRYGPK--------IVLP----QRWREFDSRGRRRDSYFYFKRK 330

Qy        439 IPKYVPPHLSPDKKWLGTPIEEMRRM-PRCGIRLPLLRPSANHTVTIRVDLLRAGEVPKP 497
               :   |:   :   :           :     |        |           :||  |
Db        331 LDGYLKCYKTTGYFMFVGLLHNMWEFDPDITYKLPALE---------------------- 368

Qy        498 FPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI-QTALLNKFTRPQNLKDA 556
                  :||::                                   ||:  :    :|  ||  |      :
Db        369 --MYYKEM-----------------------------SELVGREEVLEKFARVARIA-- 394

Qy        557 ILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKIALCLPNICTQPIPLLK 616
                                                |   |:    :  :|: :: |:
Db        395 ----------------------KTAEDILPERIYR-LVGDV------------------ 412

Qy        617 QKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDINFNRLFEGRSSRKPEKL 676
                     | |:|  :|  |:|:|   ||             |:     |  :|         | |:            |||
Db        413 ----ESATLSHKQCAALVARMFF-------ARPDSPFS------FCRILSSDKSICVEKL 455

Qy        677 KTLFCYFRRVTEKKPTGLVTF--TRQSLEDF-PEWERCEKPLTRLHVT--YEGTIEENGQ 731
               | ||  ||  :::     |  | |:|    |:      :   |     ||:    :|  |     |              ::    :  |:
Db        456 KFLFTYFDKMSMDPPDGAVSFRLTKMDKDTFNEEWK--DKKLRSLPEVEFFDEMLIEDTA 513

Qy        732 GMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNECLIITGTEQYSE 791
                      |||||    :||||   :   |  |||||||:  ||:::      |     |   :      |  :  |     |      :|
Db        514 LCTQVDFANEHLGGGVLNHGSVQEEIRFLMCPEMMVGMLLCEKMKQLEAISIVGAYVFSS 573

Qy        792 YTGYAETYRW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFRR-----YLDQFVPEK 840
               ||||    |  :|              :   |   :  :  :|   ||        |    :||||:  |:                :|
Db        574 YTGYGHTLKWAELQPNHSRQNTNEFRDRFGRLRVETIAIDAILFKGSKLDCQTEQLNKAN 633

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
               :  ||:  ||    ||:    |         |:    :  ||   ||||||  ||     ||   :||::||     |:|    :    :
Db        634 IIREMKKASIGFMSQGPKFTNI-PIVTGWWGCGAFNGDKPLKFIIQVIAAGVADRPLHFC 692

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVG 926
               :||:  ||          :       :  ::  :|:|
Db        693 SFGEPELAAKCKKIIERMKQKDVTLG 718
```

Search completed: May 26, 2004, 18:49:25
Job time : 36.2745 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 26, 2004, 18:46:21 ; Search time 81.4317 Seconds
                                         (without alignments)
                                         3343.018 Million cell updates/sec

Title:          US-09-302-812-4
Perfect score:  5190
Sequence:       1 MNAGPGCEPCTKATRWGAAT..........YHAVESCAETADHSGQRTGT 976

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters:      1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5190	100.0	976	9	US-09-973-451-4	Sequence 4, Appli
2	4635.5	89.3	977	9	US-09-973-451-2	Sequence 2, Appli
3	4395	84.7	968	9	US-09-973-451-6	Sequence 6, Appli
4	1078	20.8	768	9	US-09-973-451-8	Sequence 8, Appli
5	595	11.5	546	12	US-10-425-114-60000	Sequence 60000, A
6	463	8.9	726	9	US-09-973-451-10	Sequence 10, Appl
7	217.5	4.2	200	12	US-10-424-599-184988	Sequence 184988,
8	188	3.6	180	12	US-10-424-599-156445	Sequence 156445,
9	179.5	3.5	300	12	US-10-424-599-233915	Sequence 233915,
10	173.5	3.3	748	9	US-09-864-761-43244	Sequence 43244, A
11	167	3.2	31	9	US-09-973-451-11	Sequence 11, Appl
12	166.5	3.2	1163	12	US-10-221-278-197	Sequence 197, App
13	166.5	3.2	1163	15	US-10-291-172-197	Sequence 197, App
14	165.5	3.2	1274	12	US-10-282-122A-43837	Sequence 43837, A
15	164.5	3.2	1111	9	US-09-815-242-12955	Sequence 12955, A
16	161	3.1	2476	10	US-09-824-574-7	Sequence 7, Appli
17	159	3.1	1462	14	US-10-287-218-17	Sequence 17, Appl
18	158.5	3.1	1948	14	US-10-032-585-7611	Sequence 7611, Ap
19	156.5	3.0	1297	12	US-10-282-122A-71692	Sequence 71692, A
20	155.5	3.0	1163	10	US-09-932-257A-3	Sequence 3, Appli
21	155.5	3.0	1974	15	US-10-369-493-6395	Sequence 6395, Ap
22	154.5	3.0	498	12	US-10-424-599-196154	Sequence 196154,
23	153	2.9	6642	15	US-10-369-493-5013	Sequence 5013, Ap
24	150.5	2.9	408	12	US-10-424-599-224071	Sequence 224071,
25	150.5	2.9	1169	12	US-10-282-122A-71179	Sequence 71179, A
26	150.5	2.9	1312	15	US-10-369-493-1950	Sequence 1950, Ap
27	150.5	2.9	1957	15	US-10-369-493-2070	Sequence 2070, Ap
28	150	2.9	407	12	US-10-282-122A-71177	Sequence 71177, A
29	150	2.9	1501	9	US-09-924-154-17	Sequence 17, Appl
30	149.5	2.9	691	15	US-10-108-260A-4345	Sequence 4345, Ap
31	149.5	2.9	1210	14	US-10-205-823-264	Sequence 264, App
32	149	2.9	26	9	US-09-973-451-20	Sequence 20, Appl
33	149	2.9	29	9	US-09-973-451-12	Sequence 12, Appl
34	148.5	2.9	665	10	US-09-820-843A-107	Sequence 107, App
35	148.5	2.9	677	14	US-10-060-036-4552	Sequence 4552, Ap
36	148.5	2.9	677	15	US-10-295-027-1306	Sequence 1306, Ap
37	148.5	2.9	1928	15	US-10-369-493-22025	Sequence 22025, A
38	147.5	2.8	677	14	US-10-060-036-180	Sequence 180, App
39	146.5	2.8	903	12	US-10-282-122A-52328	Sequence 52328, A
40	146.5	2.8	1109	12	US-10-425-114-72939	Sequence 72939, A
41	146.5	2.8	2125	9	US-09-919-172-29	Sequence 29, Appl
42	146.5	2.8	2649	14	US-10-205-219-169	Sequence 169, App
43	146.5	2.8	2649	15	US-10-341-434-220	Sequence 220, App
44	146.5	2.8	2649	15	US-10-341-434-230	Sequence 230, App
45	145.5	2.8	1908	16	US-10-197-824-15	Sequence 15, Appl

ALIGNMENTS

```
RESULT 1
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
;   LENGTH: 976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-973-451-4

  Query Match             100.0%;  Score 5190;  DB 9;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  976;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNAGPGCEPCTKATRWGAATTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVPG 60

Qy         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         61 QAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFYQ 120

Qy        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVPE 180

Qy        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240

Qy        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESEP 300

Qy        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360

Qy        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPKA 420

Qy        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 EDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSANH 480

Qy        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQ 540

Qy        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVKI 600

Qy        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDIN 660

Qy        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 FNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLHV 720

Qy        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNEC 780

Qy        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPEK 840

Qy        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 MRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVYF 900

Qy        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHAV 960

Qy        961 ESCAETADHSGQRTGT 976
              ||||||||||||||||
Db        961 ESCAETADHSGQRTGT 976
```

RESULT 2
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-973-451-2

```
  Query Match             89.3%;  Score 4635.5;  DB 9;  Length 977;
  Best Local Similarity   89.5%;  Pred. No. 0;
  Matches  874;  Conservative   32;  Mismatches   69;  Indels    1;  Gaps    1;

Qy          1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
              |:||||||||||  || || || |||||||||| ||||||| ||| |||||||||  |  
Db          1 MSAGPGCEPCTKRPRWDAAATSPPAASDARSFPGRQRRVLDSKDAPVQFRVPPSSSGCAL 60

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
              |:||||||||||||||||||||||||||||| ||||| |||||||| |||||||||||||
Db         61 GRAGQHRGSATSLVFKQKTITSWMDTKGIKTVESESLHSKENNNTREESMMSSVQKDNFY 120

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
              |||:||| |||||  |||  || |||| |||||||||||||| |:|:||||||  |||||
Db        121 QHNMEKLENVSQLGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
              |||||||:|:|   |||||| ||| |||||||| ||||||||| |||  |||:||||| |
Db        181 EQFSNANVDQSSPKDDHSDTNSEESRDNQQFLTHVKLANAKQTMEDEQGREARSHQKCGK 240

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
              :||| | || |||:| |||  ||||| |||||||| |   :| |||| ||||||||||||
Db        241 ACHPAEACAGCQQEETDVVSESPLSDTGSEDVGTGLKNANRLNRQESSLGNSPPFEKESE 300

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
              |||||||||||||||||||||||||||||||| || |||||||||| |:|| | ||| | 
Db        301 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDSSSAQTANKPSRFQPREADTELRKRSSA 360

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
              ||||:|||||||||||| |||| ||| ||  |||||||||||||:|||||||| ||:||
Db        361 KGGEIRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGRKDSKITDHFMRVPK 420

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
              |||:|||| | ||||||||||||:||||||||||||||||||||||||||||| ||||||
Db        421 AEDKRKEQCEMKHQRTERKIPKYIPPHLSPDKKWLGTPIEEMRRMPRCGIRLPPLRPSAN 480

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
              ||||||||||| ||||||||||:|||||||||||||||||||||||||||| ||||||||
Db        481 HTVTIRVDLLRIGEVPKPFPTHFKDLWDNKHVKMPCSEQNLYPVEDENGERAAGSRWELI 540

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
              ||||||: ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 600

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 660

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        661 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKLLTRLH 720

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
              |||||||| ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        721 VTYEGTIEGNGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLFTEVLDHNE 780
```

```
Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        781 CLIITGTEQYSEYTGYAETYRWARSHEDRSERDDWQRRTTEIVAIDALHFRRYLDQFVPE 840

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        841 KIRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAVAERDVVY 900

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        901 FTFGDSELMRDIYSMHTFLTERKLTVGEVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 960

Qy        960 VESCAETADHSGQRTG 975
Db        961 VESCTQTTNQPGQRTG 976
```

```
RESULT 3
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
;   LENGTH: 968
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
US-09-973-451-6

  Query Match             84.7%;  Score 4395;  DB 9;  Length 968;
  Best Local Similarity   85.7%;  Pred. No. 0;
  Matches  837;  Conservative   52;  Mismatches   78;  Indels   10;  Gaps    6;

Qy          1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDPKDAHVQFRVPPSSPACVP 59
Db          1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPPSSPACVS 59

Qy         60 GQAGQHRGSATSLVFKQKTITSWMDTKGIKTAESESLDSKENNNTRIESMMSSVQKDNFY 119
Db         60 GRAGPHRGNATSFVFKQKTITTWMDTKGPKTAESE---SKENNNTRIDSMMSSVQKDNFY 116

Qy        120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQLLESEPQTVTLVP 179
Db        117 PHKVEKLENVPQLNLDKSPTEKSSQYLNQQQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176

Qy        180 EQFSNANIDRSPQNDDHSDTDSEENRDNQQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
Db        177 KQLSNANIGQSPHTDDHSDTDHEEDRDNQQFLTPIKLANTKPTVGDGQAR---SNCKCSG 233

Qy        240 SCHPGEDCASCQQDEIDVVPKSPLSDVGSEDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299
Db        234 SRQSVKDCTGCQQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPFEKESE 293

Qy        300 PESPMDVDNSKNSCQDSEADEETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYST 359
Db        294 PESPMDVDNSRNSCQDSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT 352

Qy        360 KGGEVRLHFQFEGGESRTGMNDLNAKLPGNISSLNVECRNSKQHGKKDSKITDHLMRLPK 419
Db        353 KGSEVRLHFQFE-GENNAGTSDLNAKPSGNSSSLNVECRSSKQHGKRDSKITDHFMRISK 411

Qy        420 AEDRRKEQWETKHQRTERKIPKYVPPHLSPDKKWLGTPIEEMRRMPRCGIRLPLLRPSAN 479
Db        412 SEDRRKEQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRKMPRCGIHLPSLRPSAS 471

Qy        480 HTVTIRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 539
Db        472 HTVTVRVDLLRAGEVPKPFPTHYKDLWDNKHVKMPCSEQNLYPVEDENGERTAGSRWELI 531

Qy        540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEEAEAQHLYQSILPDMVK 599
Db        532 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALVDFWDKVLEEAEAQHLYQSILPDMVK 591

Qy        600 IALCLPNICTQPIPLLKQKMNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
Db        592 IALCLPNICTQPIPLLKQKMNHSVTMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 651

Qy        660 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 719
Db        652 NFNRLFEGRSSRKPEKLKTLFCYFRRVTEKKPTGLVTFTRQSLEDFPEWERCEKPLTRLH 711

Qy        720 VTYEGTIEENGQGMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIISRLFTEVLDHNE 779
Db        712 VTYEGTIEGNGRGMLQVDFANRFVGGGVTGAGLVQEEIRFLINPELIVSRLFTEVLDHNE 771

Qy        780 CLIITGTEQYSEYTGYAETYRWSRSHEDGSERDDCERRCTEIVAIDALHFRRYLDQFVPE 839
Db        772 CLIITGTEQYSEYTGYAETYRWARSHEDGSEKDDWQRRCTEIVAIDALHFRRYLDQFVPE 831

Qy        840 KMRRELNKAYCGFLRPGVSSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 899
Db        832 KVRRELNKAYCGFLRPGVPSENLSAVATGNWGCGAFGGDARLKALIQILAAAAAERDVVY 891

Qy        900 FTFGDSELMRDIYSMHIFLTERKLTVGDVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 959
Db        892 FTFGDSELMRDIYSMHTFLTERKLDVGKVYKLLLRYYNEECRNCSTPGPDIKLYPFIYHA 951

Qy        960 VESCAETADHSGQRTGT 976
Db        952 VESSAETTDMPGQKAGT 968
```

```
RESULT 4
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
;  APPLICANT: JACOBSON, Myron K.
;  APPLICANT:  JACOBSON, Elaine L.
;  APPLICANT:  AM, Jean-Christophe
;  APPLICANT:  LIN, Winston
;  TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
;  TITLE OF INVENTION:  (PARG) ENZYMES,
;  TITLE OF INVENTION:  THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
;  TITLE OF INVENTION:  THEREWITH
;  FILE REFERENCE: NIAD 201
;  CURRENT APPLICATION NUMBER: US/09/973,451
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: US/09/302,812
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: 60/083,768
;  PRIOR FILING DATE: 1998-05-01
;  NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
;   LENGTH: 768
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
US-09-973-451-8
```